NOVEL EXPRESSED GENES

This application claims the benefit of U.S. Serial No. 10/112,372, filed April 1, 2002, U.S. Serial No. 60/382, 614, filed May 24, 2002, U.S. Serial No. 10/164,717, filed June 10, 2002, U.S. Serial No. 10/167,631, filed June 13, 2002, U.S. Serial No. 10/177,917, filed June 24, 2002, and U.S. Serial No. 60/399,125, filed 30 July 2002, which are hereby incorporated by reference in their entirety.

DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the expression of OTB0949 in human tissues. PCR was performed using SEQ NO 3 as the forward primer, and SEQ ID NO 4 as the reverse primer.

Fig. 2 shows the expression pattern of human OTB182, an integral membrane protein, in human tissues. To detect gene expression, PCR was carried out on aliquots of the normalized tissue samples using a forward (SEQ ID NO 13) and reverse (SEQ ID NO 14) gene-specific primers.

Fig. 3 shows the amino acid alignment of human OTB182 (SEQ ID NO 12) with mouse AK003645 (SEQ ID 15).

Fig. 4 (A-D) shows the amino acid alignments of a human transient receptor potential cation channel (TRPCC) with related channel family members. Human sequences are TRPM (SEQ ID NO 17), human AB046836 (SEQ ID NO 19), human XM_036123 (SEQ ID NO 18), and mouse XM_140575 (SEQ ID 20).

Fig. 5 shows the expression pattern of human TRPCC in human tissues. To detect gene expression, PCR was carried out on aliquots of the normalized tissue samples using a forward (SEQ ID NO 21) and reverse (SEQ ID NO 22) gene-specific primers.

Fig. 6 shows the amino acid sequence alignments between different forms of the human melanocortin-1 receptor. NM_002386 or MCR-1A (SEQ ID NO 30). MCR-1C (SEQ ID NO 26). MCR-1B (SEQ ID NO 31).

Fig. 7 shows a schematic of the exon sizes for the melanocortin-1 gene and the tubulin gene (exon 7).

Fig. 8 shows the expression pattern of OTB860 in human tissues. SEQ ID NOS 40 and 41 are the primer sequences.

Fig. 9 (A-C) shows the amino acid alignments of OTB860 (SEQ ID NO 39) and KIAA1678 (SEQ ID NO 42).

Fig. 10 (A and B) is the amino acid alignments of the different splice variants of human TARPP, Br137A (SEQ ID NO 46), B (SEQ ID NO 48), C (SEQ ID NO 50), D (SEQ ID NO 52; SEQ ID NO 14, NM_016300), and E (SEQ ID NO 44), and partial clone AL133109 (SEQ ID NO 55).

Fig. 11 is a schematic drawing showing the differences between the various forms of human TARPP.

Fig. 12 (A-C) shows amino acid alignments of the different splice variants of human TARPP (Br137A, B, C, D, and E) with mouse TARPP (NM_033264; SEQ ID NO 53).

The following procedure was used for the expression profile. A twenty-four tissue panel was used (lanes from left to right): 1, adrenal gland; 2, bone marrow; 3, brain; 4, colon; 5, heart; 6, intestine; 7, pancreas; 8, liver; 9, lung; 10, lymph node; 11, lymphocytes; 12, mammary gland; 13, muscle; 14, ovary; 15, pancreas; 16, pituitary; 17, prostate; 18, skin; 19, spleen; 20, stomach; 21, testis; 22, thymus; 23, thyroid; 24, uterus. The lane at the far left of each panel contains molecular weight standards. The results were obtained according to the following procedures:

Polyadenylated mRNA was isolated from tissue samples, and used as a template for first-strand cDNA synthesis. The resulting cDNA samples were normalized using beta-actin as a standard. For the normalization procedure, PCR was performed on aliquots of the first-strand cDNA using beta-actin specific primers. The PCR products were visualized on an ethidium bromide stained agarose gel to estimate the quantity of beta-actin cDNA present in each sample. Based on these estimates, each sample was diluted with buffer until each contained the same quantity of beta-actin cDNA per unit volume.

To detect gene expression, PCR was carried out on aliquots of the normalized tissue samples using a forward and reverse gene-specific primers. The reaction products were loaded on to an agarose (e.g., 1.5-2%) gel and separated electrophoretically.

DESCRIPTION OF THE INVENTION

The present invention relates to all facets of the novel genes described herein, polypeptides encoded by them, antibodies and specific binding partners thereto, and their applications to research, diagnosis, drug discovery, therapy, clinical medicine, forensic science and medicine, etc. The polynucleotides and polypeptides are useful in variety of

ways, including, but not limited to, as molecular markers, as drug targets, and for detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating, determining predisposition to, etc., diseases and conditions, associated with genes of the present invention. The identification of specific genes, and groups of genes, expressed in pathways physiologically relevant to particular tissues, permits the definition of functional and disease pathways, and the delineation of targets in these pathways which are useful in diagnostic, therapeutic, and clinical applications. The present invention also relates to methods of using the polynucleotides and related products (proteins, antibodies, etc.) in business and computer-related methods, e.g., advertising, displaying, offering, selling, etc., such products for sale, commercial use, licensing, etc.

OTB0949

OTB0949 is a polynucleotide (SEQ ID NO 1-2) which is expressed predominantly in brain tissue. Low levels of expression are observed in other tissues, e.g., adrenal gland, mammary gland, pituitary, stomach, and testes, but brain expression is at least 100-fold higher. See, e.g., Fig. 1. Because of its selectivity for brain, OTB0949 can be used as a molecular marker for brain tissue, e.g., in pathology and cytology, as well as a target, e.g., to ablate brain tissue, to deliver drugs to brain cells, etc. In the brain, OTB0949 is highly expressed in amygdala, hippocampus, thalamus, and retina. OTB0949 can also be a useful in diagnostics and therapeutics to treat neurological and visual disorders.

The brain is one of the most complicated and least understood organs in the mammalian body. Anatomically, it is composed of four different regions: (1) cerebral hemispheres, (2) diencephalon (thalamus, hypothalamus, and epithalmus), (3) brain stem (midbrain, pons, and medulla oblongata), and (4) cerebellum. These can be further subdivided. For instance, the cerebral hemispheres contain cerebral cortex and basal ganglia (caudate nucleus, putamen, globus pallidus, lentiform nucleus, corpus striatum, amygdala). The midbrain contains, e.g., cerebral peduncles, corpora quadrigemina, colliculi, substantia nigra, and the red nucleus. Other regions and subdivisions of interest include hypothalamus, pituitary, cranial nerves, pineal, gray matter, white matter, raphe nucleus, limbic system, etc. Various cell types are found in the brain, including, supporting cells, such as neuroglia, glia, astrocytes, microglia, ependymal cells, oligodendrocytes, and Schwann cells, neurons, such

as multipolar, bipolar, unipolar, Purkinje, and pyramidal cells.

The gene is located at chromosomal position 12q24.2. Several neurological diseases have been mapped to this region, e.g., spinocerebellar ataxia 2 (associated with a mutation in the ataxin-2 gene), spinal muscular atrophy (OMIM 158590), and amyotrophy (OMIM 181405). Disruption of OTB0949 (e.g., in the corresponding gene a transgenic animal) can result in a brain disorder or susceptibility thereto, including those mentioned above.

OTB0949 is coded for in a single exon, and comprises a short coding sequence of about 135 amino acids (SEQ ID NO 2). It contains a stretch of hydrophobic amino acids from about positions 76-100 (SEQ ID NO 2). Examples of promoters include, e.g., SEQ ID NOS 5-10 located in contig NT_009775 at about 557-607, 1263-1313, 1591-1641, 1635-1685, 1714-1764, and 1936-1986, respectively.

The present invention also relates to polypeptides of OTB0949, e.g., an isolated human OTB0949 polypeptide comprising or having the amino acid sequence set forth in SEQ ID NO 2, an isolated human OTB0949 polypeptide comprising an amino acid sequence having 80, 85, 90, 95, 97, 99% or more amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO 1, and optionally having one or more of OTB0949 activities, such as cell signaling activity, secretory pathway activity, etc. Fragments specific to OTB0949 can also used, e.g., to produce antibodies or other immune responses, as competitors to any of its activities, etc. These fragments can be referred to as being "specific for" OTB0949. The latter phrase, as already defined, indicates that the peptides are characteristic of OTB0949, and that the defined sequences are substantially absent from all other protein types. Such polypeptides can be of any size which is necessary to confer specificity, e.g., 5, 8, 10, 12, 15, 20, etc.

OTB182

OTB182 is an integral membrane protein comprising 307 amino acids (SEQ ID NO 12). Using SMART (e.g., Schultz et al., *Proc. Natl. Acad. Sci.*, 95:5857-5864, 1998; Letunic et al., *Nucleic Acid Res.*, 30:242-244, 2002), the protein is predicted to have seven membrane spanning regions at about amino acids positions 45-67, 87-109, 129-151, 161-183, 196-218, 231-253, and 278-297. There is a putative signal sequence (at amino acids 1-26) at its N-terminus which also overlaps with an eight transmembrane spanning domain (at amino acids

10-32).

OTB182 is expressed predominantly in excitable tissues, e.g., brain, heart, and muscle, with very low expression observed in prostate tissues. See. Fig. 2. In the brain, it is expressed predominantly in thalamus. It also expressed in neural stem cells. Its expression in excitable tissues makes OTB182 a highly selective marker for excitable tissues, as well as indicating a functional and/or developmental role in this tissue type. There are a number of genes whose expression is restricted predominantly to excitable cells, e.g., GABA-interacting factor-1 (GRIF-1; Beck et al., *J. Biol. Chem.*, 28 May 2002); calcium calmodulin dependent (CaM) kinase (e.g., Loseth et al., *Brain Research*, 869(1-2):137-145, 2000); sodium channel types (e.g., Schaller et al., *J. Neurosci*, 12(4):1370-81, 1992); calcium dependent mitochondrial solute carrier (e.g., Del Arco et al., *J. Biol. Chem.*, 273(36):23327-23334, 1998).

A mouse homolog of OTB182 is AK003645 (SEQ ID NO 15) that codes for a 153 amino acid polypeptide. It shares about 92% sequence identity from amino acids 1-122, and sharply diverges from that point onward. See. Fig. 3. Murine OTB182 maps to chromosomal location 11E2. The present invention relates all transcripts associated with the AK003645 gene loci. The degree of nucleotide sequence identity between human and mouse (AK003645) is about 83% from about nucleotide position 71-437 of SEQ ID NO 12.

OTB182 is located at chromosomal band 17q25. A genetically-inherited neuromuscular disease, hereditary neuralgic amyotrophy (HNA) has been mapped to this locus. See, e.g., Jeannet et al., *Neurology*, 57:1963-1968, 2001. In addition, a hereditary hearing loss maps to 17q25 (e.g., DFNA20, Morell et al., *Genomics*, 63:1-6, 2000) and mental retardation (Rio et al., *Human Genetics*, 108:511-515, 2001). Examples of specific polynucleotides are SEQ ID NOS 13 and 14.

Diseases or disorders which can be treated in accordance with the present invention include, but are not limited to neuropathy, neuralgic amyotrophy (e.g., HNA), myopathy, sensorineural hearing loss (e.g., DFNA20), mental retardation, neuromuscular disorders, brain cancer, such as a neuroblastoma, and other diseases and conditions involving heart, brain, and muscle tissues, etc.

A transgenic animal with OTB182 functionally disrupted can show a defect in an excitable cell. Such defect, includes, e.g., developmental defects, defects in the functional

activity of the cell, e.g., in excitability, membrane conductance, response to stimuli, signal transduction, or any of the disorders mentioned herein.

Antibodies to OTB182 can also be produced, e.g., an antibody which is specific-for: an epitope selected from amino acids 123-307 of SEQ ID NO 12, or comprising amino acid 27, 47, 64, 66, 75, 78, 105, 111, or 113 of SEQ ID NO 12

Human transient receptor potential cation channel (TRPCC) gene and polypeptide

Human TRPCC codes for a polypeptide of 1707 amino acids. As shown in Fig. 5, it is selectively expressed in brain, kidney, and pituitary, with very low expression observed in testis and ovary. By the phrase "selectively expressed," it is meant that a nucleic acid molecule, when produced as a transcript, is characteristic of the tissue or cell-type in which it is made. This can mean that the transcript is expressed only in that tissue and in no other tissue-type, or it can mean that the transcript is expressed preferentially, differentially, predominantly, and more abundantly (e.g., at least 5-fold, 10-fold, etc., or more) in that tissue when compared to other tissue-types.

The nucleotide and amino acid sequences of human TRPCC are shown in SEQ ID NOS 16 and 17. Analysis of its primary structure indicates the presence of six transmembrane domains at about amino acids 870-892, 901-1112, 904-921, 936-958, 971-990, 1005-1024, 1085-1107 of SEQ ID NO 17, however, by analogy to other ion channels, it is generally believed to have only six transmembrane spanning regions. See, e.g., Clapham et al., *Nature Reviews, Neuroscience*, 2:387, 2001. The ion transport domain comprises amino acids 901-1112. There is also a putative transmembrane domain at the N-terminus at about amino acids 5-24. According to the six-transmembrane model, both the N- and C-terminus of the protein are intracellular, and provide a scaffolding for interaction with other proteins.

The human TRPCC contains 25 exons. The present invention relates to any isolated introns and exons that are present in the gene. Intron and exon boundaries can be routinely determined, e.g., using the sequences disclosed herein.

Partial sequences for human TRPCC were previously identified (e.g., Accession Numbers AB046836 and XM_036123). For example, human AB046836 (SEQ ID 19) is incomplete, coding for 1017 amino acids (See Fig. 4, AB046836), and lacks the first 690 amino acids of human TRPCC, but shares about 99% identity with TRPCC along the rest of

its length. Another partial sequence, human XM_036123 (SEQ ID NO 18) codes for 988 amino acids (See Fig 4, XM_036123), lacking the first 719 amino acids of human TRPCC, but shares 100% identity with TRPCC along the rest of its length (See Fig 4). XM_140575 (SEQ ID NO 20) appears to be a homolog of human TRPCC, and shares about 94% sequence identity from about amino acids 82-693, or about amino acids 345-956 of human TRPCC (SEQ ID NO 17). Amino acids 1-81 and 694-736 (see Fig. 4) of the mouse homolog have low sequence identity with human TRPCC. Alignment with mouse genomic DNA using Spidey (NCBI) indicates that amino acids 1-80 of XM_140575 are derived from exons 1 and 2 of the genomic DNA, and amino acids 694-736 are derived from exon 7 of the mouse genomic DNA. XM_140575 is located on mouse chromosome 19B.

TRPCC maps to chromosomal region 9q21.1. Strikingly, hypomagnesemia with hypocalcemia (OMIM 602014) are known to be determined by a mutation within 9q21 (Walder et al., *Human Molecular Genetics*, 6: 1491-1497, 1997), as would be expected with a channel responsible for cation conductance. Consistent with its expression in brain, a susceptibility to amyotrophic lateral sclerosis with frontotemporal dementia (OMIM 105550) was mapped to this same chromosomal locus (Pinsky et al., *Clinical Genetics*, 7:186-191, 1975; Hosler et al., *JAMA*, 284:1664-1669, 2000). In addition, schizophrenia (Hovatta et al., *Am. J. Hum. Genet.*, 65:1114-24), and familial dyskinesia/facial myokymia (Fernandez et al., *Ann. Neurol.*, 49:486-92, 2001) are also associated with this gene locus. Nucleic acids of the present invention can be used, e.g., as linkage markers, diagnostic targets, and therapeutic targets for any of the mentioned disorders, as well as any disorders or genes mapping in proximity of TRPCC.

TRCC polynucleotides, polypeptides, ligands, and binding partners thereto, can be used in a number of useful ways. For example, binding partners, such as antibodies and ligands, can be used to selectively target agents to brain, kidney, and other tissues in which it is expressed for purposes including, but not limited to, imaging, diagnostic, therapeutics, etc. Imaging of tissues can be facilitated using agents such as TRPCC antibodies that can be used to target contrast agents to a specific site in the body. Various imaging techniques have been used in this context, including, e.g., X-ray, CT, CAT, MRI, ultrasound, PET, SPECT, and scintographic. A reporter agent can be conjugated or associated routinely with a TRPCC antibody. Ultrasound contrast agents combined with ligands such as antibodies are described

in, e.g., U.S. Pat. Nos 6,264,917; 6,254,852; 6,245,318; and 6,139,819. MRI contrast agents, such as metal chelators, radionucleotides, paramagnetic ions, etc., combined with selective targeting agents are also described in the literature, e.g., in U.S. Pat. Nos. 6,280,706 and 6,221,334. The methods described therein can be used generally to associate TRPCC and ligands thereof with an agent for any desired purpose.

An active agent can be associated in any manner with an TRPCC ligand that is effective to achieve its delivery to the target. The association of the active agent and the ligand ("coupling") can be direct, e.g., through chemical bonds between the binding ligand and the agent or via a linking agent, or the association can be less direct, e.g., where the active agent is in a liposome, or other carrier, and the ligand is associated with the liposome surface. In such case, the ligand can be oriented in such a way that it is able to bind to TRPCC on the surfaces of kidney or brain cells.

Useful human TRPCC polypeptides and corresponding nucleic acids include polypeptides comprising amino acids 1-88, 5-24, 1-690, 1-719, and fragments thereof (See SEQ ID NO 17 and Fig. 4). Nucleic acids and polypeptides can be used as probes (e.g., in PCR, in Northern blots, etc.), as diagnostic agents, to generate antibodies, as vaccines, to produce recombinant proteins, as antisense, etc. A specific polynucleotide according to the present invention can be determined routinely. Examples are specific probes are SEQ ID NOS 21-24, e.g., where SEQ ID NOS 23 and 24 can be used as forward and reverse PCR primers, respectively, to amplify a portion of amino acid region 1-160 of SEQ ID NO 17.

TRPCC has a number of biological activities, including, e.g., cation transport, signal transduction, protein binding, etc. By "signal transduction" is meant the activation of a chain of events that alters the concentration of one or more small intracellular signaling molecules (second messengers), e.g., cyclic AMP, calcium ions, as described in Sabala et al., *British Journal of Pharmacology,* 132:393-402, 2001. By "cation transport" is meant the influx or efflux of a cation, e.g., calcium, magnesium, into or from a cell. Mizuno et al., *Molecular Brain Research,* 64:41-51, 1999. Protein binding indicates the ability of the protein to interact with other proteins, e.g., as the N-terminus interacts with intracellular proteins. These activities can be determined routinely. Signal transduction can be assessed by expression of TRPCC in cells, etc., and measurement of the concentrations of elicited second messengers or byproducts, e.g., Ca^{2+} or Mg^{2+} or cAMP, inositol, etc., by, e.g., atomic

absorption spectrometry (ThermoElemental SOLAAR AA spectrometers), radioimmunoassay, etc. Sano et al. *Science*, 293:1327-1330, 2001. Cation transport can be assessed by measurement of changes in ionic currents by whole-cell patch-clamp analysis. For instance, cells or oocytes can be transfected with a polynucleotide of the present invention and then analyzed for expression of calcium channel activity, e.g., using patch clamp, calcium activated dyes, etc.. See, also, e.g., Strubing et al., *Neuron*, 29:645-655, 2001; Sano et al., *Science*, 293:1327, 2001; Ohki et al., *J. Biol. Chem.*, 275:39055-39060, 2000; Boulay et al., *J. Biol. Chem.*, 272:29672-29680, 1997.

The present invention relates to an isolated polynucleotide comprising, e.g., a polynucleotide sequence coding without interruption for a human TRPCC polypeptide, or complement thereto, said TRPCC having 80%, 85%, 90%, 92%, 95%, 99%, or more amino acid sequence identity along its entire length to the sequence comprising amino acids 1-690 of SEQ ID NO 17, and 80%, 85%, 90%, 92%, 95%, 99%, or more amino acid sequence identity along its entire length to the sequence comprising from amino acids 691-1707 of SEQ ID NO 17, and which has, e.g., cation transport, signal transduction, or protein binding activity.

Antibodies can be prepared against specific epitopes or domains of TRPCC, e.g., amino acids 2-30, 773-789, 870-887, 905-913, 943-958, 969-986, 1005-1022, 1087-1114, 1125-1131, 789-870, 913-943, 986-1005, etc.

Detection can be desirable for a variety of different purposes, including research, diagnostic, prognostic, and forensic. Diagnostic purposes included testing patients and their families for the presence of mutations associated with hypomagnesemia with hypocalcemia or amyotrophic lateral sclerosis with frontotemporal dementia. The selected mutant alleles, SNPs, polymorphisms, etc., can be used diagnostically to determine whether a subject has, or is susceptible to a disorder associated with TRPCC, as well as to design therapies and predict the outcome of the disorder. Methods involve, e.g., diagnosing a disorder associated with TRPCC or determining susceptibility to a disorder, e.g., hypomagnesemia with hypocalcemia or amyotrophic lateral sclerosis with frontotemporal dementia, comprising, detecting the presence of a mutation in a TRPCC gene (such as a mutation in SEQ ID NO 16, or variants thereof. The sequences of TRPCC genes can also be compared, e.g., between a normal gene as shown in SEQ ID NO 16 and the sequence of a gene from a patient with the disorder, e.g.,

hypomagnesemia with hypocalcemia.

Fragments specific to TRPCC can also be used, e.g., to produce antibodies or other immune responses, as competitors to nucleotide binding, ligand binding, etc. or as, e.g., inhibitors or stimuli in signal transduction pathways. These fragments can be referred to as being "specific for" TRPCC. The latter phrase, as already defined, indicates that the peptides are characteristic of TRPCC, and that the defined sequences are substantially absent from all other protein types. Such polypeptides can be of any size necessary to confer specificity, e.g., 5, 8, 10, 12, 15, 20, etc. Examples of polypeptides include but are not limited to polypeptides that comprise the following amino acid residues: 2-60, 598-660 of SEQ ID NO 17, or fragments thereof.

Biological activities of TRPCC include, e.g., cation channel activity, signal transduction activity, and protein binding activity. As discussed above, the biological activity of TRPCC can be measured routinely. For example, if agents are to be identified which modulate the channel activity of TRPCC either electrophysiology or calcium imaging can be used to assess their effects, e.g., using fluo-3, Fura-red, Ca-sensitive chemi-luminescent probes, etc. (e.g., kits are commercially available from Molecular Probes) and a laser scanning confocal microscope to visualize the changes in intracellular calcium as a result of modulation of TRPCC.

A transgenic animal, or animal cell, lacking one or more functional TRPCC genes can be useful in a variety of applications, including, as an animal model for hypomagnesemia with secondary hypocalcemia, amyotrophic lateral sclerosis with frontotemporal dementia, etc., drug screening assays (e.g., for signal transduction mediated by agents other than TRPCC; by making a cell deficient in TRPCC, the contribution of other receptors to, e.g., Ca^{2+} modulation can be specifically examined), as a source of tissues deficient in TRPCC activity, etc. Such an animal can show a defect in cation (e.g., calcium) conductance, e.g., an impairment in the permeation of an ion through the channel.

An isolated polynucleotide can comprise, e.g., a polynucleotide sequence coding without interruption for a human TRPCC polypeptide, or complement thereto, said TRPCC having 90% or more amino acid sequence identity along its entire length to the sequence comprising amino acids 1-690 of SEQ ID NO 17, and 90% or more amino acid sequence identity along its entire length to the sequence comprising from amino acids 691-1707 of

SEQ ID NO 17, and which has cation transport activity.

The present invention also relates to a methods of identifying a mutation associated with amyotrophic lateral sclerosis with frontotemporal dementia, comprising: comparing the structure of: genomic DNA comprising all or part of human TRPCC, mRNA comprising all or part of human TRPCC, cDNA comprising all or part of human TRPCC, or a polypeptide comprising all or part of human TRPCC, with the complete structure of human TRPCC as set forth in SEQ ID NO 16, in a patient having amyotrophic lateral sclerosis with frontotemporal dementia, or a family member thereof.

Melanocortin

The present invention relates to novel forms of a melanocortin-1 receptor (also known as "MCR-1" or alpha-melanocyte stimulating hormone receptor). It is highly expressed in melanocytes, and is a key component of the pathway which modulates skin and hair pigmentation. Moreover, certain alleles of MCR-1 are associated with a high risk of melanoma. MCR-1 is also expressed in other tissues, including monocytes, mast cells, placenta, pituitary, and endothelial cells.

MCR-1 belongs to the G-protein coupled receptor (GPCR) super-family. Its expression is restricted to melanocytes and few other cell types, such as monocytes, mast cells, and endothelial cells. See, e.g., Smith et al., *Gene*, 281:81-94, 2001; Scholzen et al., *Annals of the New York Academy of Sciences*, 885:239-253 (1999). Stimulation of the receptor by its natural ligands (e.g., alpha-melanocyte stimulating hormone or "α-MSH") causes an increase in cAMP levels which, in turn, stimulates intracellular tyrosinase activity. Increased activity of the tyrosinase enzyme drives the conversion of phaeomelanin (yellow and red pigments) to eumelalanin (brown and black pigments).

The MCR-1 gene is located at chromosomal position 16q24. It is adjacent to the tubulin TUBB4 gene, and its 3' region overlaps with the tubulin promoter (Smith et al.). Transcripts containing genic material from both MCR1 and TUBB4 have been identified, including transcipts which contain coding sequences from both. See, e.g., NCBI accession number BC020171. These may be involved in cancer.

Almost 40 different polymorphisms in the MCR receptor have been identified. See, Sturm et al., *Gene*, 277:49-62, 2001; Table 1. Several of these (e.g., Arg151Cys; Arg160Trp;

Asp294His) are strongly associated with red hair, fair skin, and poor tanning ability. It has been reported that these alleles are nonfunctional receptors and do not stimulate cAMP production when stimulated by MSH. See, Table 2. As a consequence, phaeomelanin is not converted to eumelalanin, and skin and hair color reflect the cell's high content of the yellow and red phaeomelanin pigments. Significantly, individuals who have these alleles are also at a higher risk for skin cancers, such as basal cell carcinoma, squamous cell carcinoma, and melanoma. See, e.g., Sturm et al., *Am. J. Hum. Genet.*, 6 (supplement to volume 67): 16, Oct. 2000. See, also OMIM, No. 155555 for other information on MCR-1, including disease information, polymorphisms, etc.

The present invention relates to novel forms of MCR-1. In one embodiment, the present invention relates to a novel MCR-1 variant, MCR-1 type C or MCR-1C, which possesses a unique carboxy-terminus. Previous reports had identified a 317 amino acid form of MCR-1 ("MCR-1A") in a number of different species, including human (SEQ ID NO 30), chimpanzee, muskox, sheep, cow, horse, dog, and fox. This form was characterized as full-length. Several minor size variants were observed, as well, e.g., in mouse (315 amino acid acids), in pig (310 amino acids), and in cow (321 amino acids). A second form, MCR-1B, was (SEQ ID NO 31) also reported that had an additional 65 amino acids at its terminus (Tan et al., *FEBS Letters*, 451:137-141, 1991; WO 00/39147). The present invention relates to a third form of MCR-1 (MCR-1C) that comprises 32 carboxy-amino acids (amino acids 367-398 of SEQ ID NO 26) not previously identified in any melanocortin receptor variant. This novel form comprises part of the new carboxy terminus identified in MCR-1B, but diverges from it at amino acid position 367. See, Fig. 6.

Fig. 7 shows exons which have been detected in melanocortin-1 receptors. Exons 1, 2, and 3 contain MCR coding sequences; Exons 5, 6, and 7 contain tubulin coding sequences.

MCR-1A (e.g., NM_002386: SEQ ID NO 30) contains exon 1, and MCR-1B contains exons 1 and 2. MCR-1C contains coding sequence from exons 1-3. As indicated by the stop codon TGA, exon 3 comprises both coding and noncoding sequence. MCR-1C can also contain noncoding sequences, e.g., exons 4, 5, 6, and/or 7 (e.g., for a total of exons 1-7). BC020171, mentioned above, contains the coding sequence from exons 1 and 2 fused to the coding sequences of tubulin in exons 5-7. Examples of promoters for MCR-1C include, e.g., SEQ ID NOS 35-37.

The present invention also relates to a polymorphism at amino acid position 120, where an isoleucine (I) is replaced with a threonine (T). Isoleucine is present at amino acid position 120 in most melanocortin receptor-1 homologs, except pig which has a methionine substitution. This polymorphism may affect the receptor's functionality. Analysis of the transmembrane structure using TMHMM v. 2.0 (Krogh et al., *Journal of Molecular Biology*, 305(3):567-580, January 2001; Sonnhammer et al., In J. Glasgow et al., editors, *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*, pages 175-182, Menlo Park, CA, 1998. AAAI Press; Moller et al., *Bioinformatics*, 17(7):646-653, July 2001) reveals a different number of predicted transmembrane sequences than the isoleucine isoform.

The present invention relates to any polynucleotide, or polypeptide encoded thereby, which codes for MCR-1C, including receptors having any polymorphism, such as the naturally-occurring polymorphisms listed in Tables 1 and 2, and those disclosed herein. Examples include SEQ ID 27 having 120T and 163Q, SEQ 28 having 120I and 163R, and SEQ ID 29 having 120I and 163Q. It also includes polynucleotide and polypeptide fragments which are specific for MCR-1C (e.g., 367-398 of SEQ ID NO 26 and fragments thereof), and polynucleotides and polypeptides which comprise such specific fragments. For example, the present invention relates to a polynucleotide comprising a coding sequence for amino acid 367-398 of SEQ ID NO 26, or fragments thereof, such as any five amino acid sequence contained therein.

The present invention also relates to an isolated polynucleotide comprising, a polynucleotide sequence coding without interruption for a human MCR-1C, said MCR-1C having about 80%, 85%, 88%, 89%, 90%, 92%, 95%, 99%, etc., or more amino acid sequence identity along its entire length to the amino acid sequence set forth in SEQ ID NO 26, or a complement thereto, and which has ligand-binding activity, G-protein binding activity, or cAMP production activity. For example, such a polynucleotide can comprise one or more of the polymorphisms listed in Tables 1 and 2 (e.g., if 36 of the listed polymorphisms were present in such a polynucleotide, it would have about 90% (360/398) sequence identity along its entire length to the amino acid sequence of SEQ ID NO 26. A corresponding amount of nucleotide is included, e.g., 90%, 92%, 95%, 97%, 98%, 99%, or more.

Similarly, the present invention relates to an isolated polynucleotide comprising, a polynucleotide sequence coding without interruption for a human MCR-1C, or complement thereto, said MCR-1C having 80%, 84%, 85%, 86%, 88%, 90%, 95%, or more amino acid sequence identity along its entire length to the sequence comprising amino acids 1-316 of SEQ ID NO 26, and 85%, 90%, 92%, 95%, etc. or more amino acid sequence identity along its entire length to the sequence comprising from amino acids 317-398 of SEQ ID NO 26, and which has ligand-binding activity, G-protein binding activity, or cAMP production activity.

As stated, a polynucleotide can code for a polypeptide having one or more of the following activities, ligand-binding activity, G-protein binding activity, cAMP production activity, or other functional activities. Ligand-binding activity indicates the ability of MCR-1C to bind specifically to a receptor ligand, such as a ACTH, MSH, etc. Ligand binding activity can be using a radioactive or otherwise labeled ligand, or whole-cell assays using labeled ligands. See, e.g., WO0039147, WO9957148, U.S. Pat. Nos. 5,731,408, 6,100,048, and 6,350,760; Libert et al., *Pigment Cell Res.*, 2:510-518, 1989.

G-protein binding activity indicates the ability of the receptor protein to bind to a G-protein. Such binding does not determined routinely, e.g., using filtration assays necessarily have to be productive, i.e., the binding does not have to result in stimulation of the signal transduction cascade. G-protein binding can be measured using in vivo and in vitro binding assays, as well as functional assays. See, e.g., Ford et al., *Science*, 280:1271-1274, 1998.

cAMP production is a measure of the ability of the receptor to stimulate the generation of cAMP upon binding by a receptor agonist. MCR is known to couple to G-proteins and thereby activate adenyl cyclase, increasing intracellular levels of cAMP (e.g., Buckley & Ramachandran, *Proc. Natl. Acad. Sci.*, 78: 7431-7435, 1981; Grahame-Smith et al., 1967, *J. Biol. Chem.* 242: 5535-5541; Mertz & Catt, 1991, *Proc. Natl. Acad. Sci.* 88: 8525-8529; Pawalek et al., 1976, *Invest. Dermatol.* 66: 200-209). This property of cells expressing the MCR-1C can be used assess its "cAMP production activity." For example, cells can be transfected with MCR-1C DNA, plated, and washed once with DMEM containing 1% bovine serum albumin (BSA) and 0.5 mM IBMX (a phosphodiesterase inhibitor). The cells can then be treated with hormone (e.g., alpha-MSH, gamma-MSH, ACTH, etc.). Following hormone treatment, the cells can be washed with phosphate buffered

saline, and intracellular cAMP extracted by lysing the cells. Intracellular cAMP concentrations can be determined routinely, e.g., using an assay (Amersham) which measures the ability of cAMP to displace cAMP from a high affinity cAMP binding protein (see Gilman, 1970, *Proc. Natl. Acad. Sci.*, 67: 305-312).

Polynucleotide and polypeptides of the present invention can be used for a variety of purposes, including, but not limited to, treating cancers, treating skin cancer and other cancers modulating skin and hair pigmentation, identifying MCR ligands, modulating the MCR-1 receptor types, determining susceptibility to skin cancer, detecting MCR-1C expression, determining polymorphisms in MCR-1C, making MCR-1C polypeptide, expressing MCR-1C in host cells, making antibodies to MCR-1 receptor types, modulating cutaneous inflammation (see, e.g., Bhardwaj et al., *J. Immunol.*, 158:3378-3384, 1997; Luger et al., *Ann. NY Acad. Sci.*, 917:232-238, 2000), modulating melanocytes, monocytes, endothelial cells, or other cells in which MCR-1C is expressed, etc.

The expression of MCR-1C on the surface of melanoma cells makes it a useful target. Melanoma is a skin cancer which originates from melanocytes present normally in the epidermis and underlying cell layers. There are four basic types: lentigo maligma melanoma, superficial spreading melanoma, nodular melanoma, and acral lentigous melanoma. Because of its expression on melanocytes, MCR-1C specific antibodies and other binding partners can be used to treat melanoma, e.g., by conjugating cytotoxic agents to antibodies directed to the receptor. In addition, MCR-1C polynucleotides, polypeptides, and binding partners thereto can be used to detect metastatic melanoma cells.

Modulation of the MCR-1C can also be used to modulate skin pigmentation, e.g., to increase the amount brown and black pigments to darken skin color, to provide protective effects against UV radiation, to block receptor activation, e.g., preventing or reducing the accumulation of brown and black pigments in the skin, preventing or reducing tanning, preventing or reducing skin freckling, etc. Agonists and antagonists of the melanocortin receptor, include, alpha-melanocyte stimulating hormone and adrenocorticotropic hormone. Other ligands are disclosed in, e.g., WO9957148, U.S. Pat. Nos. 5,731,408, 6,100,048, and 6,350,760, and can be identified and isolated as described in these patents, as well as WO0039147.

As discussed earlier, several MCR-1 alleles have been associated with a greater risk

of skin cancer. For example, the presence of the Asp84Glu variant imposed a high risk of melanoma in individual carriers. See, Kennedy et al., *J. Invest. Dermatol.*, 117:294-300, 2001. Other alleles with increased risk of melanoma included, Val60Leu, Val92Met, Arg142His, Arg151Cys, Arg160Trp, Arg163Gln, and His260Pro (Kennedy et al.). See, also, Scott et al., *J. Cell. Sci.*, 115 (Pt. 11):2349-2355, 2002. MCR-1C of the present invention can be used to assess melanoma risks, e.g., determining the presence of a variant of MCR-1C in individuals, and whether such variants are associated with skin cancer and other melanocyte disorders. Analysis can be performed by any suitable method, e.g., by single-stranded conformation polymorphism analysis and DNA sequence analysis.

Expression can also be "selective," where expression is observed. By the phrase "selectively expressed," it is meant that a nucleic acid molecule comprising the defined sequence of nucleotides, when produced as a transcript, is characteristic of the tissue or cell-type in which it is made. This can mean that the transcript is expressed only in that tissue and in no other tissue-type, or it can mean that the transcript is expressed preferentially, differentially, and more abundantly (e.g., at least 5-fold, 10-fold, etc., or more) in that tissue when compared to other tissue-types.

In view of their selectivity and display on the cell surface, MCR-1C polypeptides of the present invention are a useful target for histological, diagnostic, and therapeutic applications relating to the cells in which they are expressed. Antibodies and other protein binding partners (e.g., ligands, aptamers, small peptides, etc.) can be used to selectively target agents to a tissue for any purpose, included, but not limited to, imaging, therapeutic, diagnostic, drug delivery, gene therapy, etc. For example, binding partners, such as antibodies, can be used to treat melanomas in analogy to how c-erbB-2 antibodies are used to breast cancer. They can also be used to detect metastatic cells, in biopsies, etc. The genes and polypeptides encoded thereby can also be used in tissue engineering to identify tissues as they appear during the differentiation process, to target tissues, to modulate tissue growth (e.g., from starting stem cell populations), etc. Useful antibodies or other binding partners include those that are specific for parts of the polypeptide which are exposed extracellularly. Any of the methods described above and below can be accomplished in vivo, in vitro, or ex vivo.

Binding partners can also be used as to specifically deliver therapeutic agents to a

tissue of interest. For example, a gene to be delivered to a tissue can be conjugated to a binding partner (directly or through a polymer, etc.), in liposomes comprising cell surface, and then administered as appropriate to the subject who is to be treated. Additionally, cytotoxic, cytostatic, and other therapeutic agents can be delivered specifically to the tissue to treat and/or prevent any of the conditions associated with the tissue of interest.

The present invention relates to methods of detecting melanoma cells, comprising one or more of the following steps, e.g., contacting a sample comprising cells with a polynucleotide specific for MCR-1C (e.g., amino acids 367-398, and fragments thereof), or a mammalian homolog thereof, under conditions effective for said polynucleotide to hybridize specifically to said gene, and detecting specific hybridization. Detecting can be accomplished by any suitable method and technology, including, e.g., any of those mentioned and discussed below, such as Northern blot and PCR. Specific polynucleotides include SEQ ID NOS 32-34, and complements thereto.

As indicated above, binding partners can be used to deliver agents specifically to melanocytes, e.g., for diagnostic, therapeutic, and prognostic purposes, including the treatment of melanoma. Methods of delivering an agent to a melanocyte cell can comprise, e.g., contacting a melanocyte with an agent coupled to binding partner specific for a melanocortin receptor gene of the present invention, whereby said agent is delivered to said cell. Any type of agent can be used, including, therapeutic and imaging agents. Contact with the melanocyte (e.g., a melanoma) can be achieved in any effective manner, including by administering effective amounts of the agent to a host orally, parentally, locally, systemically, intravenously, etc. The phrase "an agent coupled to binding partner" indicates that the agent is associated with the binding partner in such a manner that it can be carried specifically to the target site. Coupling includes, chemical bonding, covalent bonding, noncovalent bonding (where such bonding is sufficient to carry the agent to the target), present in a liposome or in a lipid membrane, associated with a carrier, such as a polymeric carrier, etc. The agent can be directly linked to the binding partner, or via chemical linkers or spacers.

Imaging of specific organs can be facilitated using tissue selective antibodies and other binding partners that selectively target contrast agents to a specific site in the body. Various imaging techniques have been used in this context, including, e.g., X-ray, CT, CAT, MRI, ultrasound, PET, SPECT, and scintographic. A reporter agent can be conjugated or

associated routinely with a binding partner. Ultrasound contrast agents combined with binding partners, such as antibodies, are described in, e.g., U.S. Pat. Nos. 6,264,917, 6,254,852, 6,245,318, and 6,139,819. MRI contrast agents, such as metal chelators, radionucleotides, paramagnetic ions, etc., combined with selective targeting agents are also described in the literature, e.g., in U.S. Pat. Nos. 6,280,706 and 6,221,334. The methods described therein can be used generally to associate a partner with an agent for any desired purpose.

A transgenic animal with a disrupted melanocortin-1C receptor can have a pigmentation phenotype, e.g., red or fair hair. Functional disruption of the gene can be accomplished in any effective way, including, e.g., introduction of a stop codon into any part of the coding sequence, e.g., to prevent expression of amino acids 367-398, such that the resulting polypeptide is biologically inactive or lacks one or more of its functional regions, introduction of a mutation into a promoter or other regulatory sequence that is effective to turn it off, or reduce transcription of the gene, insertion of an exogenous sequence into the gene which inactivates it (e.g., which disrupts the production of a biologically-active polypeptide or which disrupts the promoter or other transcriptional machinery), deletion of sequences from the gene, etc. A transgenic animal, or animal cell, lacking one or more functional genes of the present invention can be useful in a variety of applications, including, as an animal model for conditions and diseases associated with melanocortin-1C, for drug screening (e.g., by making a cell deficient in MCR-1C, the contribution of the activity remaining variants, such as MCR-1B and the 317-amino acid form, can be assessed), as a source of tissues deficient in one or more MCR-1 activities. The animal's endogenous locus can be replaced with a continuous coding sequence for MCR-1C, such that only MCR-1C is expressed, and no other form, such as MCR-1B and the 317-amino acid form, are expressed.

PCR based methods can also be used in the methods of detecting polynucleotides for human MCR-1C. In such methods, more than one probe specific for MCR-1C can be used, e.g., a pair of specific polynucleotide probes which are capable of amplifying a polynucleotide sequence of MCR-1C, such as corresponding to amino acids 1-366, 367-398, etc., of SEQ ID NO 26. For instance, SEQ ID NO 32 is in exon 1, SEQ ID NO 33 spans exons 2-3, and SEQ ID NO 34 is in exon 4. Thus, in a PCR reaction, SEQ IDS 32 and 33 produce a fragment about 262 base pairs that is absent in MCR-1A and MCR-1B. SEQ ID

NOS 32 and 34 in a PCR reaction produce a fragment of about 615 base pairs which is absent from MCR-1A and MCR-1B.

Mutant alleles, polymorphisms, SNPs, etc., can be identified and isolated from melanomas and other skin conditions that are known, or suspected to have, a genetic component. Identification of such genes can be carried out routinely (see, above for more guidance), e.g., using PCR, hybridization techniques, direct sequencing, mismatch reactions (see, e.g., above), RFLP analysis, SSCP (e.g., Orita et al., *Proc. Natl. Acad. Sci.*, 86:2766, 1992), etc., where a polynucleotide having a sequence selected from SEQ ID NO 25 (especially corresponding to amino acids 367-398) can be used as a probe. The selected mutant alleles, SNPs, polymorphisms, etc., can be used diagnostically to determine whether a subject has, or is susceptible to a melanoma or other condition (e.g., pigmentation variation, inflammatory condition) associated with a melanocortin receptor gene of the present invention, as well as to design therapies and predict the outcome of the disorder.

The present invention an isolated polynucleotide comprising, a polynucleotide sequence coding without interruption for a human MCR-1C, or complement thereto, said MCR-1C having 84% or more amino acid sequence identity along its entire length to the sequence comprising amino acids 1-316 of SEQ ID NO 26, and 90% or more amino acid sequence identity along its entire length to the sequence comprising from amino acids 317-398 of SEQ ID NO 26, and which has ligand-binding activity, G-protein binding activity, or cAMP production activity.

OTB860

OTB860 codes for an intracellular polypeptide comprising 1700 amino acids. Its nucleotide and amino acid sequences are shown in SEQ ID NOS 38 and 39. Expression of OTB860 is detected predominantly in heart and brain tissues, with minimally detectable levels in breast and testes tissues. The expression pattern is illustrated in Fig. 8.

As shown in Fig. 9, OTB860 is related to KIAA1678 (also known as AB051465; SEQ ID NO 42). It contains 369 amino acids at its N-terminal end which are not present in KIAA1678, and a 29 amino acid insertion (1546-1574) at about amino acid position 1545. It also differs at amino acid positions 847 (histidine instead of glutamine) and 867 (arginine instead of glutamine).

Consistent with the expression pattern of OTB860, there are a number of functional and developmental pathways that are shared between neuronal and cardiac cells. For example, neuregulins (neuregulin-1 – NRG1) play an important role in myocardial and neuronal development. Mice deficient in IgL-domain containing neuregulins have severe defects in the developing heart and nervous system. See, e.g., Kramer et al., *Proc. Natl. Acad. Sci.*, 93: 4833-4838, 1996; Zhao et al., *J. Biol. Chem.*, 273:10261-10269, 1998. These effects appear to be mediated by Type I neuregulins. Meyer et al., *Develop*, 124:3575-3586, 1997. The neuroregulin receptor, erbB4, is also highly expressed in heart and brain, although its expression is not restricted to these tissues (data not shown). Adhesion pathways are also shared between the two tissues. For example, alpha4-integrin is expressed in both neuronal and cardiac cells. See, e.g., Pinco et al., *Mech. Dev.*, 100:99-103, 2001.

OTB860 maps to chromosomal band 2q36. A number of disorders have been mapped to, or in close proximity to, this chromosome location. These include, e.g., brachydactyly type A1, pili torti and nerve deafness, syndactyly type 1, Gracile syndrome (growth retardation and early death), and epilepsy. Recently, in a whole genome scan, 2q36 was identified as a locus associated with acute coronary syndrome (Harrap et al., *Arterioscler. Thromb. Vasc. Biol.*, 22:874-878, 2002), involving myocardial infarction, unstable angina, atherosclerotic plaque disruption, and coronary thrombosis. Nucleic acids of the present invention can be used as linkage markers, diagnostic targets, therapeutic targets, for any of the mentioned disorders, as well as any disorders or genes mapping in proximity to it.

OTB860 can be used in diagnostic, therapeutic, prophylactic, and research applications. RNA and polypeptide detection methods can be used to determine whether a sample comprises neuronal or cardiac tissues. When a positive is obtained, cell type markers can be used to determine precisely whether the tissue is neuronal or cardiac. For example, the presence or absence of a neuronal marker would distinguish between brain and heart tissues. Non-limiting examples of neuronal markers include, presenilins, genes and polypeptides in the pathways for neurotransmitter synthesis, receptor, metabolism, etc., (e.g., serotonin, MAO, dopamine, norephinephrine, nitric oxide, etc.), apolipoprotein A, APP, neuron-specific enolase (NSE), glial fibrillary acidic protein (GFAP), S100, GAP-43, neuron-specific beta-III tubulin, Stac (neuron-specific protein with an SH3 domain, e.g., Genomics, 47:140-2, 1998), myelin basic protein, vimentin, etc. Non-limiting examples of heart tissue

markers include, cardiac troponin I, and smooth muscle markers such as CRP1.

The present invention also relates to polypeptide detection methods for assessing heart or brain function, e.g., methods of assessing heart or brain function, comprising, detecting OTB860 polypeptide, or fragments thereof, in a body fluid, whereby the level of OTB860 polypeptide in said fluid is a measure of heart or brain function. Heart or brain function tests are usually performed to determine whether the organ is functioning normally as a way of diagnosing disease. Various tests are used to test for heart function, such as electrocardiogram, stress test, echocardiogram, oxygen levels, and cardiac enzyme tests (e.g., creatine phosphokinase, troponin, lactate dehydrogenase, and myoglobin). Detection of OTB860 provides an additional assessment tool, especially in diseases such as myocardial infarction and other conditions, e.g., where cellular debris, etc., is released systemically. As with the other tests, elevated levels of OTB860 in blood, or other fluids, can indicate impaired brain or heart function. Values can be determined routinely, as they are for other functional markers.

OTB860 polynucleotides and polypeptides can be used to treat, prevent and diagnose diseases of the heart, including, e.g., acute coronary syndrome, myocardial hypertrophy, heart failure, conduction disordres, arrhythmias, bradyarrhythmias, sinus node dysfunction, tachyarrhythmias, tachycardias, atrial fibrillation, congenital heart diseases (see, e.g., Harrison's *Principles of Internal Medicine*, Volume 1, 12th Edition, 1991, Pages 924-925), atrial and ventricular septal defects, congenital aortic stenosis, coartation of the aorta, valvular heart disease, myocardial infarction, ischemic heart disease, cardiomyopathy, perocardial diseases, cardiac tumors (e.g., myxoma, lpoma, papillary fibroelastoma, rhabdomyoma, sarcoma, etc), coronary artery disease, atherosclerosis, aortic aneurysm, etc.

OTB860 polynucleotides and polypeptides can also be used to treat, prevent and diagnose diseases of the brain, including, e.g., vascular diseases, hypoxia, ischemia, infarction, tumors, neuroglial tumors, astrocytoma, glioblastoma multiforne, pilocytic astrocytoma, oligodendroglioma, ependymona, choriod plexus papilloma, neuronal tumors, neuroblastoma, ganglioneuroma, gangliocytoma, gangliogioma, primitive or undifferentiated tumors, medulloblastoma, tumors of meninges, mingioma, lymphomas, demyelinating diseases, multiple sclerosism perivenous encephalitis, degenerative diseases, Alzheimer's, Pick's, Huntington's. Parkinsonism, ALS, Werdnig-Hoffman, degenerative diseases of

cerebral cortex, ganglia, brainstem, and motor neurons, inborn errors of metabolism, demyelinating and dysmyelinating disorders, Pelizaeus-Merzbacher disease, multiple sclerosis, various leukodystrophies, post-traumatic demyelination, cerebrovascular (CVS) accidents, neuritis, neuropathies, particularly, multifocal leucoencephalopathy, Guillain-Barre syndrome, retrobulbar neuritis, acute rubella encephalitis, chronic relapsing polyneuropathy, intravascular lymphomatosis, Krabbe disease, globoid cell leukodystrophy, subacute combined degeneration of the spinal cord and brain, allergic encephalitis, murine caronavirus, hepatitis virus infection, or Theiler's murine encephalomyelitis, prion diseases, Creutzfeldt-Jakob, especially, febrile familial convulsions, epilepsy, vascular neuromyopathy, cerebellar ataxia, etc.

When expression is described as being "predominantly" in a given tissue, this indicates that the gene's mRNAs levels are highest in this tissue as compared to the other tissues in which it was measured. Expression can also be "selective," where expression is observed. By the phrase "selectively expressed," it is meant that a nucleic acid molecule comprising the defined sequence of nucleotides, when produced as a transcript, is characteristic of the tissue or cell-type in which it is made. This can mean that the transcript is expressed only in that tissue and in no other tissue-type, or it can mean that the transcript is expressed preferentially, differentially, and more abundantly (e.g., at least 5-fold, 10-fold, etc., or more) in that tissue when compared to other tissue-types.

The present invention relates to methods of detecting brain or heart cells, comprising one or more of the following steps, e.g., contacting a sample comprising cells with a polynucleotide specific for OTB860 (e.g., SEQ ID NOS 40-41), or a mammalian homolog thereof, under conditions effective for said polynucleotide to hybridize specifically to said gene, and detecting specific hybridization. Detecting can be accomplished by any suitable method and technology, including, e.g., any of those mentioned and discussed below, such as Northern blot and PCR. Specific polynucleotides include SEQ ID NOS 40-41, and complements thereto.

Detection can also be achieved using binding partners, such as antibodies (e.g., monoclonal or polyclonal antibodies) that specifically recognize polypeptides coded for by genes of the present invention. Thus, the present invention relates to methods of detecting a brain or heart cell, comprising, one or more the following steps, e.g. contacting a sample

comprising cells with a binding partner (e.g. an antibody, an Fab fragment, a single-chain antibody, an aptamer) specific for a polypeptide coded for by OTB860 (e.g., SEQ ID NO 39), or a mammalian homolog thereof, under conditions effective for said binding partner bind specifically to said polypeptide, and detecting specific binding. Protein binding assays can be accomplished routinely, e.g., using immunocytochemistry, ELISA format, Western blots, etc. Useful epitopes include those exposed to the surface.

A brain or heart cell (see above for examples of brain or heart cell types) can also be modulated in accordance with the present invention, e.g., by methods of modulating a brain or heart cell, comprising, e.g., contacting said cell with an agent effective to modulate OTB860, or the biological activity of a polypeptide encoded thereby (e.g., SEQ ID NO 39), or a mammalian homolog thereof, whereby said brain or heart cell is modulated. Modulation as used throughout includes, e.g., stimulating, increasing, agonizing, activating, amplifying, blocking, inhibiting, reducing, antagonizing, preventing, decreasing, diminishing, etc. Any activity or function of the brain or heart cell can be modulated, including, e.g., development, differentiation, signaling, excitability, etc.

The present invention also relates to methods of modulating development of cardiac or neuronal cells, comprising, e.g., administering an agent which is effective for modulating the expression of OTB860, or the biological activity of a polypeptide encoded thereby, whereby the development of said cardiac or neuronal cell is modulated. Development is meant to include any process in which a cell or tissue matures, including differentiation, organogenesis, cell proliferation, cell survival, expression and induction of functional molecules, cell movement and migration, apoptosis, modulation of gene expression, trabeculation, etc. Examples of heart and brain development that can be modulation are disclosed in Kramer et al., *Proc. Natl. Acad. Sci.*, 93:4833-4838, 1996. Any agent can be used to modulate development in any environment, e.g., in situ, in vivo, or in vitro.

OTB860 can be used to detect, modulate, etc., any of the cell types present in the heart or brain, including, but not limited to, heart cells comprising the coverings of the heart (e.g., pericardium, fibrous pericardium, serous pericardium containing the parietal layer and epicardium), heart wall comprising mycardium, cardiac muscle, and endocardium (endothelial), blood vessels, valves, and autorhythmic cardiac cells such as those in the SA and VA nodes, brain and other neuronal cells, such neurons, glia, microglia, ependymal cells,

oligodendrocytes, Schwann cells, and satellite cells.

Promoter sequences obtained from OTB860 can be utilized to selectively express heterologous genes in brain or heart cells. Methods of expressing a heterologous polynucleotide in brain or heart cells can comprise, e.g., expressing a nucleic acid construct in brain or heart cells, said construct comprising a promoter sequence operably linked to said heterologous polynucleotide, wherein said promoter sequence is obtained from OTB860, e.g., on genomic NT_022115.8. In addition to the cell lines mentioned below, the construct can be expressed in primary cells or in established cell lines.

The present invention also relates to methods of modulating development of cardiac or neuronal cells, comprising, e.g., administering an agent which is effective for modulating the expression of OTB60, or the biological activity of a polypeptide encoded thereby, whereby the development of said cardiac or neuronal cell is modulated.

The present invention also relates to a mammalian cell whose genome comprises a functional disruption of the human OTB860 gene within a polynucleotide sequence coding for amino acid residues 1-369 (SEQ ID NO 39) or 1546-1574 (SEQ ID NO 39). A non-human, transgenic mammal comprising such a cell can have a heart or neuronal tissue defect.

Antibodies can be produced, e.g., an antibody which is specific-for: an epitope comprising amino acid 847, amino acid 867, or an epitope contained with amino acids 1-369 (SEQ ID NO 39), or amino acids 1546-1574 (SEQ ID NO 39).

TARRP

Human TARPP (thymocyte cyclic AMP regulated phosphoprotein, or, Br137A, B, C, D, and E) is represented by a family of alternative splice variants. Figs. 10-12 summarize the differences between the multiple forms. Br137E is an 847 amino acid polypeptide. Its nucleotide and amino acid sequences are shown in SEQ ID NOS 43 and 44. Br137B (SEQ ID NO 47 and 48) has a deletion of amino acids 267-300, Br137A (SEQ ID NO 45 and 46) has a deletion of amino acids 312-331, and Br137C (SEQ ID NO 49 and 50) has a deletion of both these domains. Br137D (SEQ ID NO 51 and 52) contains only the first 87 amino acids followed by a two-amino acid N-terminus which differs from the other forms. A partial clone, AL133109 (SEQ ID NO 55) as shown in Fig. 10, is missing the first 161 amino acids of Br137E, as well as having an amino acid difference at position 312 (SEQ ID NO 44).

Br137E contains a nuclear localization signal at about amino acids 107-124, an R3H domain (single-stranded nucleic acid binding domain) at about amino acids 147-224, and a proline rich region at about amino acids 476-682. These domains are also present in the A-C splice forms, but at different amino acid positions. Human TARPP has nucleic acid binding activity conferred by the corresponding binding domain indicating that it can bind nucleic acids, preferably single-stranded DNA or RNA. This binding activity can be assayed routinely, e.g., using gel electrophoresis band shift assays, e.g., as carried out in, e.g., U.S. Pat. No. 6,333,407 and 5,789,538, ELISA-based assays (e.g., Mercury™ TransFactor Kit from Clontech), and other assays which detect DNA-protein interactions.

The Br137 family represent the human homologs of murine TARPP (thymocyte ARPP) (NM_033264; SEQ ID NO 53; "Mouse" in Fig. 12). Br137E has about 83% amino acid identity and 87% homology with it (calculated using the BLAST algorithm). See, Fig. 12 (NM_033264 is murine TARPP). In addition to amino acid sequence differences between the two proteins, human TARPP has an insertion at about amino acid positions 549-572 of SEQ ID NO 44 which is not present in the mouse protein. See, Fig. 12.

Originally, a 21 kDa polypeptide was isolated from rat basal ganglia based on its phosphorylation by cAMP-dependent protein kinase (PKA). Williams et al., *J. Neurosci.*, 9:3631-3637, 1989. It was named ARPP-21 (cAMP-regulated phosphoprotein). Activation of dopamine receptors resulted in an increase in the phosphorylation of ARPP-21. Caporaso et al., *Neuropharm.*, 39:1637-1644, 2000. Human ARPP-21 (Br137D) contains 89 amino acids (NM_016300; SEQ ID NO 52).

A high molecular weight polypeptide of ARPP-21 was subsequently identified in T-cells and named TARPP. Kisielow et al., *Eur. J. Immunol.*, 31:1141-1149, 2001. This polypeptide contains ARPP-21 sequence at its 5' end, but a novel 3' end coding for more than 700 additional amino acids (for a total of 807 amino acids). Murine TARPP appears to be involved in the regulation of thymocyte maturation and TCR rearrangement. Expression of TARPP is down-regulated after the TCR signals delivered. It is highly expressed in immature thymocytes and is associated with the commitment to the T-cell lineage, making it highly selective marker for T-cell commitment. See, Kisielow, *ibid.* After commitment to the T-cell lineage during positive selection, its expression is turned off.

There appear to be several members of the human TARPP family. KIAA0029 is a

hypothetical protein that shares about 45% amino acid sequence identity and 59% homology with Br137E. KIAA1002, a second hypothetical protein, has about 42% amino acid identity and 54% homology to it.

Human TARPP is highly expressed in brain, pituitary, muscle, and thymus. It is expressed at lower levels in adrenal gland, bone marrow, heart, small intestine, kidney, liver, ovary, prostate, stomach, testis, and thyroid. There was virtually no detectable expression in colon, lung, lymph node, peripheral lymphocytes, mammary gland, pancreas, and uterus.

As indicated by its expression pattern, human TARPP is involved the maturation of T-cells, especially in the rearrangement of the TCR. For this reason, it can be used to modulate T-cells, e.g., in allergy, autoimmune disease (e.g., rheumatoid arthritis and multiple sclerosis), and graft-host disease. It can also be used a marker to determine the index of mature versus immature T-cells, where human TARPP is marker of immature T-cells. Additionally, human TARPP is phosphorylated upon dopamine receptor activation, indicating an involvement in dopamine pathways. Consequently, it is target for diseases that involve dopamine, including, e.g., schizophrenia, substance abuse and addiction, anxiety, Parkinson's disease, and other dopaminergic diseases and conditions.

Human TARPP is localized to chromosomal band 3p21.33. There are several disorders genetically mapped to this region, including, e.g., retinal vasculopathy with cerebral leukodystrophy (OMIM 192315), deafness, neurosensory, autosomal recessive 6 (OMIM 600971), and lung cancer. Nucleic acids of the present invention can be used as linkage markers, diagnostic targets, therapeutic targets, for any of the mentioned disorders, as well as any disorders or genes mapping in proximity to it.

Diseases or disorders which can be treated in accordance with the present invention include, but are not limited to autoimmune disease, such as multiple sclerosis and rheumatoid arthritis, and allergy

The gene can be disrupted in a specific region, e.g., in the sequence coding for amino acids 1-161 of a human TARPP. Cells and/or animals can also have targeted deletions, e.g., deletion of a coding sequence for amino acids 267-300 and/or 312-331 of a human TARPP. One or more the different splice forms, Br137A-E can also be knocked-out or disrupted, e.g., to dissect out the individual activities.

The present invention relates to methods of modulating T-cells, comprising, contacting T-cells with an agent which is effective for regulating a human TARPP gene expressed in said cells, or for modulating the biological activity of a polypeptide encoded thereby.

Included also in the present invention are engineered cells, e.g., a human cell whose genome comprises a functional disruption of human TARPP in the region comprising the coding sequence for amino acids 1-161 of a human TARPP of SEQ ID NO 44, or a human cell whose genome comprises a deletion of a coding sequence for amino acids 267-300 and/or 312-331 of a human TARPP of SEQ ID NO 44.

Antibodies can be produced, e.g., an antibody which is specific-for a human TARPP, said antibody which is specific for an epitope present in amino acid sequences 1-161, 88-161, 267-300, 312-331, or a polypeptide comprising amino acid 312, of a human TARPP of SEQ ID NO 44.

LAT-1

Liver-associated transmembrane protein-1 ("LAT-1" or TMD008) codes for a polypeptide comprising 276 amino acids. Its expression is highly restricted to the liver, i.e., it is predominantly expressed in the liver. The nucleotide and amino acid sequences of it are shown in SEQ ID NOS 58 and 59. It contains transmembrane domains at about amino acid positions 24-46, 59-81, 101-123, 144-166, 203-225, and 237-259. It is homologous to the olfactory class of GPCR receptors. LAT-1 is also known as XM_060456 and AX242289.

The gene for LAT-1 maps to chromosomal band 1q22. Several different disorders map to this location, including, e.g., porphyria variegata, progression of lymphoma, Zellweger syndrome, Charcot-Marie-Tooth neuropathy-1B, congenital hypomyelination, nemaline myopathy, and CD3 zeta chain deficiency, medullary thyroid carcinoma, susceptibility to Vivax malaria, schizophrenia susceptability locus, autosomal dominant deafness, susceptibility to Lupus nephritis, familial hemiplegic migraine, apolipoprotein A-II deficiency, and familial hyperlipidemia. Nucleic acids of the present invention can be used as linkage markers, diagnostic targets, therapeutic targets, for any of the mentioned disorders, as well as any disorders or genes mapping in proximity to it.

LAT-1 can be used as a diagnostic and prognostic marker for liver function and

disease, including any of the liver diseases already mentioned. For instance, blood serum levels of LAT-1 (as well as other bodily fluids) can be used as an indicator of liver disease, especially those diseases characterized by necrotic and degenerative lesions, such as hepatitis, toxicity, and cirrhosis. Any condition which results in degeneration of the liver can result in the appearance of higher than normal amounts of blood serum LAT-1. LAT-1 can be used alone, or in combination with other molecular markers for liver function, such as bilirubin, serum aminotransferases (e.g., AST and ALT), alkaline phosphatase, gamma-glutamyltranspeptidase (GGT), albumin, globulin, and blood ammonia.

Because of the selectivity of LAT-1 for the liver, it is a useful target for both histological and therapeutic applications. Antibodies and other LAT-1 binding partners can be used to selectively target agents to liver tissue for any purpose, included, but not limited to, imaging, therapeutic, diagnostic, drug delivery, gene therapy, etc. For example, LAT-1 binding partners, such as antibodies, can be used to treat liver carcinoma, in analogy to how c-erbB-2 antibodies are used to breast cancer, to detect metastatic liver cells, etc. Useful antibodies or other binding partners include those that are specific for parts of LAT-1 which are exposed extracellularly, e.g. , amino acids 1-23, 82-100, 167-202, etc.

Imaging of specific organs can be facilitated using agents, such as LAT-1, that can be used to selectively target contrast agents to a specific site in the body. Various imaging techniques have been used in this context, including, e.g., X-ray, CT, CAT, MRI, ultrasound, PET, SPECT, and scintographic. A reporter agent can be conjugated or associated routinely with a LAT-1 binding partner. Ultrasound contrast agents combined with binding partners, such as antibodies, are described in, e.g., U.S. Pat. Nos, 6,264,917, 6,254,852, 6,245,318, and 6,139,819. MRI contrast agents, such as metal chelators, radionucleotides, paramagnetic ions, etc., combined with selective targeting agents are also described in the literature, e.g., in U.S. Pat. Nos. 6,280,706 and 6,221,334. The methods described therein can be used generally to associate a LAT-1 binding partner with an agent for any desired purpose.

LAT-1 binding partners can also be used as to specifically deliver therapeutic agents to the liver. For example, hypercholesterolemia and other metabolic diseases can be treated by gene therapy, using the LAT-1 to specifically deliver the LDL receptor to the liver. The gene can be conjugated to a LAT-1 binding partner (directly or through a polymer, etc.), in liposomes comprising cell surface. Additionally, cytotoxic, cytostatic, and other therapeutic

agents can be delivered to the liver via LAT-1 to treat and/or prevent any of the above-mentioned conditions associated with liver disease, e.g., carcinoma.

The liver is the largest and most metabolically complex organ in the body. Its functions include, e.g., storage of iron, production of bile to facilitate digestion, detoxifications of various exogenous chemicals, including alcohol and many drugs, energy stockpiling (carbohydrates and fat), production of clotting factors, and manufacture of blood. There are a number of diseases which affect the liver, including, Alagille syndrome, alcoholic liver disease, alpha-1-antitrypsin deficiency, autoimmune hepatitis, Budd-Chiari syndrome, biliary atresia, Byler disease, liver cancer, Caroli disease, cirrhosis, Crigler-Najjar syndrome, Dubin-Johnson syndrome, fatty liver, galactosemia, Gilbert syndrome, glycogen storage disease, hemangioma, hemochromatosis, hepatitis A-G, porphyria cutanea tarda, primary biliary cirrhosis, protoporphyria, erythrohepatic, Rotor syndrome, sclerosing cholangitis, and Wilson disease. Liver disease is of grave concern around the world.

The liver is divided into many small units, known as lobules. The lobule is the structural unit of the liver. Each lobule is comprised of radial plates of liver cells, called hepatocytes, and is surrounded by a connective sheath. A central vein ("CV") is located in the middle, and there are portal triads at the vertices. Each triad comprises a branch of the hepatic artery (supplying arterial blood to the lobule), a branch of the hepatic portal vein (carrying nutrient-rich blood from the digestive viscera), and a bile duct. The blood from the artery and portal vein flow into leaky capillaries, the liver sinusoids, which are located between the hepatic plates of the lobule.

The acinus is the functional unit of the liver. While the boundaries of the lobule are well visible, those of the acinus are unrecognizable under the microscope. Arising like a berry, a grape (latin "acinus") on the vine around the portal triad, the liver acinus is formed of a mass of liver cells and sinusoids which drain toward two adjacent central veins. The principal metabolic functions of the liver are performed by hepatocytes. These functions include, e.g., formation and excretion of bile, regulation of carbohydrate homeostasis, lipid synthesis and secretion of plasma lipoproteins, regulation of cholesterol metabolism, formation of urea, serum albumin, clotting factors, enzymes, and numerous other proteins; and metabolism or detoxification of drugs and other foreign substances. Hepatocytes in different regions of the acinus perform different functions, e.g., gluconeogenesis is primarily

a function of the zone of cells closest to the triad, whereas glycolysis mainly occurs in the farthest zone from it.

A promoter obtained from the LAT-1 can be used, e.g., in gene therapy to obtain tissue-specific expression of a heterologous gene (e.g., coding for a therapeutic product or cytotoxin). A promoter sequence is found at about nucleotide positions 1164-1212 of SEQ ID NO 58 and can be used (e.g., 1164 to the first ATG codon) to drive liver-specific expression of a heterologous sequence. 5' and 3' sequences (including, UTRs and introns) can be used to modulate or regulate stability, transcription, and translation of nucleic acids, including the sequence to which is attached in nature, as well as heterologous nucleic acids. A polyadenylation site is found at about nucleotide positions 4265-4275 of SEQ ID NO 58. The upstream 3'UTR can be used as described above. Useful polypeptides include polypeptides exposed extracellularly, e.g. , amino acids 1-23, 82-100, 167-202, of SEQ ID NO 59, etc.

The present invention also relates to methods of detecting human liver tissue in a sample, e.g., comprising tissue, cells, or other cellular materials or debris, comprising one or more of the following steps, e.g., contacting said sample with a binding partner specific for human LAT-1 under conditions effective for said binding partner to bind specifically to human LAT-1, and detecting specific binding between said binding partner and said human LAT-1, whereby specific binding indicates that liver tissue is present in said sample.

The sample can be contacted with the binding partner in any manner which is effective to give the binding partner access to the material present in the tissue sample. How contact is achieved can depend on the format of the detection assay. For instance, if a ELISA assay is used, and the binding partner is an antibody on a solid phase in a well, then placing an aqueous sample in the well would achieve contact between partner and sample. Any type of sample can be used, including, e.g., blood (whole blood, fractionated blood, serum, etc.), stool, urine, cerebral spinal fluid, tissue biopsy, etc.

The binding partner, such as a monoclonal or polyclonal antibody, is specific for LAT-1, and is contacted with the sample under conditions effective for said binding partner to bind specifically to human LAT-1, if human LAT-1 is present in the sample. Specific binding, as previously discussed for polynucleotides, indicates that the binding partner binds or attaches to its target polypeptide without significant binding to other polypeptides ("non-

specific binding") in the sample. This concept is well known in the art. The detection of specific binding can be accomplished by any of the aforementioned assays.

The present invention also relates to polypeptide detection methods for assessing liver function, e.g., methods of assessing liver function, comprising, detecting LAT-1 polypeptide, or fragments thereof, in a body fluid, whereby the level of LAT-1 polypeptide in said fluid is a measure of liver function. Liver function tests are usually performed to determine whether the liver is functioning normally as a way of diagnosing liver disease. Various tests are commonly used, including, e.g., alkaline phosphatase, alanine transferase, aspartate transferase, bilirubin, gamma-glutamyl transpeptidase, lactic dehydrogenase, 5'-nucleotidase, albumin, alpha-fetoprotein, mitochondrial antibodies, and prothrombin time. See, e.g., *Harrison's Principles of Internal Medicine*, Volume 2, Pages 1308-1317, 12th Edition, 1991. Detection of LAT-1 provides an additional assessment tool, especially in diseases such as hepatitis, carcinoma, liver toxicity, cirrhosis, and other liver conditions, e.g., where cellular debris, etc., is released systemically. As with the other tests, elevated levels of LAT-1 in blood, or other fluids, can indicate impaired liver function. Values can be determined routinely, as they are for other liver function markers.

Nucleic acids

A mammalian polynucleotide, or fragment thereof, of the present invention is a polynucleotide having a nucleotide sequence obtainable from a natural source. When the species name is used, e.g., a human, it indicates that the polynucleotide or polypeptide is obtainable from a natural source. It therefore includes naturally-occurring normal, naturally-occurring mutant, and naturally-occurring polymorphic alleles (e.g., SNPs), differentially-spliced transcripts, splice-variants, etc. By the term "naturally-occurring," it is meant that the polynucleotide is obtainable from a natural source, e.g., animal tissue and cells, body fluids, tissue culture cells, forensic samples. Natural sources include, e.g., living cells obtained from tissues and whole organisms, tumors, cultured cell lines, including primary and immortalized cell lines. Naturally-occurring mutations can include deletions (e.g., a truncated amino- or carboxy-terminus), substitutions, inversions, or additions of nucleotide sequence. These genes can be detected and isolated by polynucleotide hybridization according to methods which one skilled in the art would know, e.g., as discussed below.

A polynucleotide according to the present invention can be obtained from a variety of different sources. It can be obtained from DNA or RNA, such as polyadenylated mRNA or total RNA, e.g., isolated from tissues, cells, or whole organism. The polynucleotide can be obtained directly from DNA or RNA, from a cDNA library, from a genomic library, etc. The polynucleotide can be obtained from a cell or tissue (e.g., from an embryonic or adult tissues) at a particular stage of development, having a desired genotype, phenotype, disease status, etc. A polynucleotide which "codes without interruption" refers to a polynucleotide having a continuous open reading frame ("ORF") as compared to an ORF which is interrupted by introns or other noncoding sequences.

Polynucleotides and polypeptides (including any part of a differentially regulated cancer gene) can be excluded as compositions from the present invention if, e.g., listed in a publicly available databases on the day this application was filed and/or disclosed in a patent application having an earlier filing or priority date than this application and/or conceived and/or reduced to practice earlier than a polynucleotide in this application.

As described herein, the phrase "an isolated polynucleotide which is SEQ ID NO," or "an isolated polynucleotide which is selected from SEQ ID NO," refers to an isolated nucleic acid molecule from which the recited sequence was derived (e.g., a cDNA derived from mRNA; cDNA derived from genomic DNA). Because of sequencing errors, typographical errors, etc., the actual naturally-occurring sequence may differ from a SEQ ID listed herein. Thus, the phrase indicates the specific molecule from which the sequence was derived, rather than a molecule having that exact recited nucleotide sequence, analogously to how a culture depository number refers to a specific cloned fragment in a cryotube.

As explained in more detail below, a polynucleotide sequence of the invention can contain the complete sequence as shown in the corresponding SEQ ID, degenerate sequences thereof, anti-sense, muteins thereof, genes comprising said sequences, full-length cDNAs comprising said sequences, complete genomic sequences, fragments thereof, homologs, primers, nucleic acid molecules which hybridize thereto, derivatives thereof , etc.

Genomic

The present invention also relates genomic DNA from which the polynucleotides of the present invention can be derived. A genomic DNA coding for a human, mouse, or other

mammalian polynucleotide, can be obtained routinely, for example, by screening a genomic library (e.g., a YAC library) with a polynucleotide of the present invention, or by searching nucleotide databases, such as GenBank and EMBL, for matches. Promoter and other regulatory regions (including both 5' and 3' regions, as well introns) can be identified upstream or downstream of coding and expressed RNAs, and assayed routinely for activity, e.g., by joining to a reporter gene (e.g., CAT, GFP, alkaline phosphatase, luciferase, galatosidase). A promoter obtained from a gene can be used, e.g., in gene therapy to obtain tissue-specific expression of a heterologous gene (e.g., coding for a therapeutic product or cytotoxin). 5' and 3' sequences (including, UTRs and introns) can be used to modulate or regulate stability, transcription, and translation of nucleic acids, including the sequence to which is attached in nature, as well as heterologous nucleic acids.

Constructs

A polynucleotide of the present invention can comprise additional polynucleotide sequences, e.g., sequences to enhance expression, detection, uptake, cataloging, tagging, etc. A polynucleotide can include only coding sequence; a coding sequence and additional non-naturally occurring or heterologous coding sequence (e.g., sequences coding for leader, signal, secretory, targeting, enzymatic, fluorescent, antibiotic resistance, and other functional or diagnostic peptides); coding sequences and non-coding sequences, e.g., untranslated sequences at either a 5' or 3' end, or dispersed in the coding sequence, e.g., introns.

A polynucleotide according to the present invention also can comprise an expression control sequence operably linked to a polynucleotide as described above. The phrase "expression control sequence" means a polynucleotide sequence that regulates expression of a polypeptide coded for by a polynucleotide to which it is functionally ("operably") linked. Expression can be regulated at the level of the mRNA or polypeptide. Thus, the expression control sequence includes mRNA-related elements and protein-related elements. Such elements include promoters, enhancers (viral or cellular), ribosome binding sequences, transcriptional terminators, etc. An expression control sequence is operably linked to a nucleotide coding sequence when the expression control sequence is positioned in such a manner to effect or achieve expression of the coding sequence. For example, when a promoter is operably linked 5' to a coding sequence, expression of the coding sequence is

driven by the promoter. Expression control sequences can include an initiation codon and additional nucleotides to place a partial nucleotide sequence of the present invention in-frame in order to produce a polypeptide (e.g., pET vectors from Promega have been designed to permit a molecule to be inserted into all three reading frames to identify the one that results in polypeptide expression). Expression control sequences can be heterologous or endogenous to the normal gene.

A polynucleotide of the present invention can also comprise nucleic acid vector sequences, e.g., for cloning, expression, amplification, selection, etc. Any effective vector can be used. A vector is, e.g., a polynucleotide molecule which can replicate autonomously in a host cell, e.g., containing an origin of replication. Vectors can be useful to perform manipulations, to propagate, and/or obtain large quantities of the recombinant molecule in a desired host. A skilled worker can select a vector depending on the purpose desired, e.g., to propagate the recombinant molecule in bacteria, yeast, insect, or mammalian cells. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBS, pD10, Phagescript, phiX174, pBK Phagemid, pNH8A, pNH16a, pNH18Z, pNH46A (Stratagene); Bluescript KS+II (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR54 0, pRIT5 (Pharmacia). Eukaryotic: PWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene), pSVK3, PBPV, PMSG, pSVL (Pharmacia), pCR2.1/TOPO, pCRII/TOPO, pCR4/TOPO, pTrcHisB, pCMV6-XL4, etc. However, any other vector, e.g., plasmids, viruses, or parts thereof, may be used as long as they are replicable and viable in the desired host. The vector can also comprise sequences which enable it to replicate in the host whose genome is to be modified.

Hybridization

Polynucleotide hybridization, as discussed in more detail below, is useful in a variety of applications, including, in gene detection methods, for identifying mutations, for making mutations, to identify homologs in the same and different species, to identify related members of the same gene family, in diagnostic and prognostic assays, in therapeutic applications (e.g., where an antisense polynucleotide is used to inhibit expression), etc.

The ability of two single-stranded polynucleotide preparations to hybridize together is a measure of their nucleotide sequence complementarity, e.g., base-pairing between

nucleotides, such as A-T, G-C, etc. The invention thus also relates to polynucleotides, and their complements, which hybridize to a polynucleotide comprising a nucleotide sequence as set forth in the sequences disclosed herein, and genomic sequences thereof. A nucleotide sequence hybridizing to the latter sequence will have a complementary polynucleotide strand, or act as a template for one in the presence of a polymerase (i.e., an appropriate polynucleotide synthesizing enzyme). The present invention includes both strands of polynucleotide, e.g., a sense strand and an anti-sense strand.

Hybridization conditions can be chosen to select polynucleotides which have a desired amount of nucleotide complementarity with the nucleotide sequences set forth in SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58and genomic sequences thereof. A polynucleotide capable of hybridizing to such sequence, preferably, possesses, e.g., about 70%, 75%, 80%, 85%, 87%, 90%, 92%, 95%, 97%, 99%, or 100% complementarity, between the sequences. The present invention particularly relates to polynucleotide sequences which hybridize to the nucleotide sequences set forth in the sequence disclosure herein or genomic sequences thereof, under low or high stringency conditions. These conditions can be used, e.g., to select corresponding homologs in non-human species.

Polynucleotides which hybridize to polynucleotides of the present invention can be selected in various ways. Filter-type blots (i.e., matrices containing polynucleotide, such as nitrocellulose), glass chips, and other matrices and substrates comprising polynucleotides (short or long) of interest, can be incubated in a prehybridization solution (e.g., 6X SSC, 0.5% SDS, 100 μg/ml denatured salmon sperm DNA, 5X Denhardt's solution, and 50% formamide), at 22-68°C, overnight, and then hybridized with a detectable polynucleotide probe under conditions appropriate to achieve the desired stringency. In general, when high homology or sequence identity is desired, a high temperature can be used (e.g., 65 °C). As the homology drops, lower washing temperatures are used. For salt concentrations, the lower the salt concentration, the higher the stringency. The length of the probe is another consideration. Very short probes (e.g., less than 100 base pairs) are washed at lower temperatures, even if the homology is high. With short probes, formamide can be omitted. See, e.g., *Current Protocols in Molecular Biology*, Chapter 6, Screening of Recombinant Libraries; Sambrook et al., *Molecular Cloning*, 1989, Chapter 9.

For instance, high stringency conditions can be achieved by incubating the blot overnight (e.g., at least 12 hours) with a polynucleotide probe in a hybridization solution containing, e.g., about 5X SSC, 0.5% SDS, 100 μg/ml denatured salmon sperm DNA and 50% formamide, at 42°C, or hybridizing at 42°C in 5X SSPE, 0.5% SDS, and 50% formamide, 100 μg/ml denatured salmon sperm DNA, and washing at 65°C in 0.1X SSC and 0.1% SDS.

Blots can be washed at high stringency conditions that allow, e.g., for less than 5% bp mismatch (e.g., wash twice in 0.1% SSC and 0.1% SDS for 30 min at 65°C), i.e., selecting sequences having 95% or greater sequence identity.

Other non-limiting examples of high stringency conditions includes a final wash at 65°C in aqueous buffer containing 30 mM NaCl and 0.5% SDS. Another example of high stringent conditions is hybridization in 7% SDS, 0.5 M $NaPO_4$, pH 7, 1 mM EDTA at 50°C, e.g., overnight, followed by one or more washes with a 1% SDS solution at 42°C.

Whereas high stringency washes can allow for, e.g., less than 10%, less than 5% mismatch, etc., reduced or low stringency conditions can permit up to 20% nucleotide mismatch. Hybridization at low stringency can be accomplished as above, but using lower formamide conditions, lower temperatures and/or lower salt concentrations, as well as longer periods of incubation time.

Hybridization can also be based on a calculation of melting temperature (Tm) of the hybrid formed between the probe and its target, as described in Sambrook et al.. Generally, the temperature Tm at which a short oligonucleotide (containing 18 nucleotides or fewer) will melt from its target sequence is given by the following equation: Tm = (number of A's and T's) x 2°C + (number of C's and G's) x 4°C. For longer molecules, Tm = 81.5 + 16.6 $\log_{10}[Na^+]$ + 0.41(%GC) - 600/N where $[Na^+]$ is the molar concentration of sodium ions, %GC is the percentage of GC base pairs in the probe, and N is the length. Hybridization can be carried out at several degrees below this temperature to ensure that the probe and target can hybridize. Mismatches can be allowed for by lowering the temperature even further.

Stringent conditions can be selected to isolate sequences, and their complements, which have, e.g., at least about 90%, 95%, or 97%, nucleotide complementarity between the and a target polynucleotide.

Other homologs of polynucleotides of the present invention can be obtained from mammalian and non-mammalian sources according to various methods. For example, hybridization with a polynucleotide can be employed to select homologs, e.g., as described in Sambrook et al., *Molecular Cloning*, Chapter 11, 1989. Such homologs can have varying amounts of nucleotide and amino acid sequence identity and similarity to such polynucleotides of the present invention. Mammalian organisms include, e.g., mice, rats, monkeys, pigs, cows, etc. Non-mammalian organisms include, e.g., vertebrates, invertebrates, zebra fish, chicken, Drosophila, C. elegans, Xenopus, yeast such as S. pombe, S. cerevisiae, roundworms, prokaryotes, plants, Arabidopsis, artemia, viruses, etc. The degree of nucleotide sequence identity between human and mouse can be about, e.g. 70% or more, 85% or more for open reading frames, etc.

Alignment

Alignments can be accomplished by using any effective algorithm. For pairwise alignments of DNA sequences, the methods described by Wilbur-Lipman (e.g., Wilbur and Lipman, *Proc. Natl. Acad. Sci.*, 80:726-730, 1983) or Martinez/Needleman-Wunsch (e.g., Martinez, *Nucleic Acid Res.*, 11:4629-4634, 1983) can be used. For instance, if the Martinez/Needleman-Wunsch DNA alignment is applied, the minimum match can be set at 9, gap penalty at 1.10, and gap length penalty at 0.33. The results can be calculated as a similarity index, equal to the sum of the matching residues divided by the sum of all residues and gap characters, and then multiplied by 100 to express as a percent. Similarity index for related genes at the nucleotide level in accordance with the present invention can be greater than 70%, 80%, 85%, 90%, 95%, 99%, or more. Pairs of protein sequences can be aligned by the Lipman-Pearson method (e.g., Lipman and Pearson, *Science*, 227:1435-1441, 1985) with k-tuple set at 2, gap penalty set at 4, and gap length penalty set at 12. Results can be expressed as percent similarity index, where related genes at the amino acid level in accordance with the present invention can be greater than 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, or more. Various commercial and free sources of alignment programs are available, e.g., MegAlign by DNA Star, BLAST (National Center for Biotechnology Information), BCM (Baylor College of Medicine) Launcher, etc. BLAST can be used to calculate amino acid sequence identity, amino acid sequence homology, and nucleotide

sequence identity. These calculations can be made along the entire length of each of the target sequences which are to be compared.

After two sequences have been aligned, a "percent sequence identity" can be determined. For these purposes, it is convenient to refer to a Reference Sequence and a Compared Sequence, where the Compared Sequence is *compared* to the Reference Sequence. Percent sequence identity can be determined according to the following formula: Percent Identity = 100 [1-(C/R)], wherein C is the number of differences between the Reference Sequence and the Compared Sequence over the length of alignment between the Reference Sequence and the Compared Sequence where (i) each base or amino acid in the Reference Sequence that does not have a corresponding aligned base or amino acid in the Compared Sequence, (ii) each gap in the Reference Sequence, (iii) each aligned base or amino acid in the Reference Sequence that is different from an aligned base or amino acid in the Compared Sequence, constitutes a difference; and R is the number of bases or amino acids in the Reference Sequence over the length of the alignment with the Compared Sequence with any gap created in the Reference Sequence also being counted as a base or amino acid. When it is stated that a polynucleotide sequence has a certain percentage ving 95% or more sequence identity along the entire length of the polynucleotide sequence set forth in SEQ ID NO 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, or 58, it means that when the polynucleotide is shorter than the mentioned SEQ ID NOS, the missing bases are counted for the purposes of the calculation. Percent sequence identity can also be determined by other conventional methods, e.g., as described in Altschul et al., *Bull. Math. Bio.* 48: 603-616, 1986 and Henikoff and Henikoff, *Proc. Natl. Acad. Sci.* USA 89:10915-10919, 1992.

Specific polynucleotide probes

A polynucleotide of the present invention can comprise any continuous nucleotide sequence of SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, sequences which share sequence identity thereto, or complements thereof. The term "probe" refers to any substance that can be used to detect, identify, isolate, etc., another substance. A polynucleotide probe is comprised of nucleic acid can be used to detect, identify, etc., other nucleic acids, such as DNA and RNA.

These polynucleotides can be of any desired size that is effective to achieve the specificity desired. For example, a probe can be from about 7 or 8 nucleotides to several thousand nucleotides, depending upon its use and purpose. For instance, a probe used as a primer PCR can be shorter than a probe used in an ordered array of polynucleotide probes. Probe sizes vary, and the invention is not limited in any way by their size, e.g., probes can be from about 7-2000 nucleotides, 7-1000, 8-700, 8-600, 8-500, 8-400, 8-300, 8-150, 8-100, 8-75, 7-50, 10-25, 14-16, at least about 8, at least about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, or more, etc. The polynucleotides can have non-naturally-occurring nucleotides, e.g., inosine, AZT, 3TC, etc. The polynucleotides can have 100% sequence identity or complementarity to a sequence of SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, or it can have mismatches or nucleotide substitutions, e.g., 1, 2, 3, 4, or 5 substitutions. The probes can be single-stranded or double-stranded.

In accordance with the present invention, a polynucleotide can be present in a kit, where the kit includes, e.g., one or more polynucleotides, a desired buffer (e.g., phosphate, tris, etc.), detection compositions, RNA or cDNA from different tissues to be used as controls, libraries, etc. The polynucleotide can be labeled or unlabeled, with radioactive or non-radioactive labels as known in the art. Kits can comprise one or more pairs of polynucleotides for amplifying nucleic acids specific for polynucleotides, e.g., comprising a forward and reverse primer effective in PCR. These include both sense and anti-sense orientations. For instance, in PCR-based methods (such as RT-PCR), a pair of primers are typically used, one having a sense sequence and the other having an antisense sequence.

Another aspect of the present invention is a nucleotide sequence that is specific to, or for, a selective polynucleotide. The phrases "specific for" or "specific to" a polynucleotide have a functional meaning that the polynucleotide can be used to identify the presence of one or more target genes in a sample and distinguish them from non-target genes. It is specific in the sense that it can be used to detect polynucleotides above background noise ("non-specific binding"). A specific sequence is a defined order of nucleotides (or amino acid sequences, if it is a polypeptide sequence) which occurs in the polynucleotide, e.g., in the nucleotide sequences of SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, and which is characteristic of that target sequence, and substantially no non-target sequences. A probe or mixture of probes can comprise a sequence or sequences that are specific to a plurality of

target sequences, e.g., where the sequence is a consensus sequence, a functional domain, etc., e.g., capable of recognizing a family of related genes. Such sequences can be used as probes in any of the methods described herein or incorporated by reference. Both sense and antisense nucleotide sequences are included. A specific polynucleotide according to the present invention can be determined routinely.

A polynucleotide comprising a specific sequence can be used as a hybridization probe to identify the presence of, e.g., human or mouse polynucleotide, in a sample comprising a mixture of polynucleotides, e.g., on a Northern blot. Hybridization can be performed under high stringent conditions (see, above) to select polynucleotides (and their complements which can contain the coding sequence) having at least 90%, 95%, 99%, etc., identity (i.e., complementarity) to the probe, but less stringent conditions can also be used. A specific polynucleotide sequence can also be fused in-frame, at either its 5' or 3' end, to various nucleotide sequences as mentioned throughout the patent, including coding sequences for enzymes, detectable markers, GFP, etc, expression control sequences, etc.

A polynucleotide probe, especially one that is specific to a polynucleotide of the present invention, can be used in gene detection and hybridization methods as already described. In one embodiment, a specific polynucleotide probe can be used to detect whether a particular tissue or cell-type is present in a target sample, e.g., with OTB182, OTB860, or LAT-1. To carry out such a method, a selective polynucleotide can be chosen which is characteristic of the desired target tissue. Such polynucleotide is preferably chosen so that it is expressed or displayed in the target tissue, but not in other tissues which are present in the sample. For instance, if detection of brain tissue is desired, it may not matter whether the selective polynucleotide is expressed in other tissues, as long as it is not expressed in cells normally present in blood, e.g., peripheral blood mononuclear cells. Starting from the selective polynucleotide, a specific polynucleotide probe can be designed which hybridizes (if hybridization is the basis of the assay) under the hybridization conditions to the selective polynucleotide, whereby the presence of the selective polynucleotide can be determined.

Probes which are specific for polynucleotides of the present invention can also be prepared using involve transcription-based systems, e.g., incorporating an RNA polymerase

promoter into a selective polynucleotide of the present invention, and then transcribing anti-sense RNA using the polynucleotide as a template. See, e.g., U.S. Pat. No. 5,545,522.

Polynucleotide composition

A polynucleotide according to the present invention can comprise, e.g., DNA, RNA, synthetic polynucleotide, peptide polynucleotide, modified nucleotides, dsDNA, ssDNA, ssRNA, dsRNA, and mixtures thereof. A polynucleotide can be single- or double-stranded, triplex, DNA:RNA, duplexes, comprise hairpins, and other secondary structures, etc. Nucleotides comprising a polynucleotide can be joined via various known linkages, e.g., ester, sulfamate, sulfamide, phosphorothioate, phosphoramidate, methylphosphonate, carbamate, etc., depending on the desired purpose, e.g., resistance to nucleases, such as RNAse H, improved in vivo stability, etc. See, e.g., U.S. Pat. No. 5,378,825. Any desired nucleotide or nucleotide analog can be incorporated, e.g., 6-mercaptoguanine, 8-oxo-guanine, etc.

Various modifications can be made to the polynucleotides, such as attaching detectable markers (avidin, biotin, radioactive elements, fluorescent tags and dyes, energy transfer labels, energy-emitting labels, binding partners, etc.) or moieties which improve hybridization, detection, and/or stability. The polynucleotides can also be attached to solid supports, e.g., nitrocellulose, magnetic or paramagnetic microspheres (e.g., as described in U.S. Pat. No. 5,411,863; U.S. Pat. No. 5,543,289; for instance, comprising ferromagnetic, supermagnetic, paramagnetic, superparamagnetic, iron oxide and polysaccharide), nylon, agarose, diazotized cellulose, latex solid microspheres, polyacrylamides, etc., according to a desired method. See, e.g., U.S. Pat. Nos. 5,470,967, 5,476,925, and 5,478,893.

Polynucleotide according to the present invention can be labeled according to any desired method. The polynucleotide can be labeled using radioactive tracers such as ^{32}P, ^{35}S, ^{3}H, or ^{14}C, to mention some commonly used tracers. The radioactive labeling can be carried out according to any method, such as, for example, terminal labeling at the 3' or 5' end using a radiolabeled nucleotide, polynucleotide kinase (with or without dephosphorylation with a phosphatase) or a ligase (depending on the end to be labeled). A non-radioactive labeling can also be used, combining a polynucleotide of the present invention with residues having immunological properties (antigens, haptens), a specific affinity for certain reagents

(ligands), properties enabling detectable enzyme reactions to be completed (enzymes or coenzymes, enzyme substrates, or other substances involved in an enzymatic reaction), or characteristic physical properties, such as fluorescence or the emission or absorption of light at a desired wavelength, etc.

Nucleic acid detection methods

Another aspect of the present invention relates to methods and processes for detecting polynucleotides. Detection methods have a variety of applications, including for diagnostic, prognostic, forensic, and research applications. To accomplish gene detection, a polynucleotide in accordance with the present invention can be used as a "probe." The term "probe" or "polynucleotide probe" has its customary meaning in the art, e.g., a polynucleotide which is effective to identify (e.g., by hybridization), when used in an appropriate process, the presence of a target polynucleotide to which it is designed. Identification can involve simply determining presence or absence, or it can be quantitative, e.g., in assessing amounts of a gene or gene transcript present in a sample. Probes can be useful in a variety of ways, such as for diagnostic purposes, to identify homologs, and to detect, quantitate, or isolate a polynucleotide of the present invention in a test sample.

Assays can be utilized which permit quantification and/or presence/absence detection of a target nucleic acid in a sample. Assays can be performed at the single-cell level, or in a sample comprising many cells, where the assay is "averaging" expression over the entire collection of cells and tissue present in the sample. Any suitable assay format can be used, including, but not limited to, e.g., Southern blot analysis, Northern blot analysis, polymerase chain reaction ("PCR") (e.g., Saiki et al., *Science*, 241:53, 1988; U.S. Pat. Nos. 4,683,195, 4,683,202, and 6,040,166; *PCR Protocols: A Guide to Methods and Applications*, Innis et al., eds., Academic Press, New York, 1990), reverse transcriptase polymerase chain reaction ("RT-PCR"), anchored PCR, rapid amplification of cDNA ends ("RACE") (e.g., Schaefer in *Gene Cloning and Analysis: Current Innovations*, Pages 99-115, 1997), ligase chain reaction ("LCR") (EP 320 308), one-sided PCR (Ohara et al., *Proc. Natl. Acad. Sci.*, 86:5673-5677, 1989), indexing methods (e.g., U.S. Pat. No. 5,508,169), *in situ* hybridization, differential display (e.g., Liang et al., *Nucl. Acid. Res.*, 21:3269-3275, 1993; U.S. Pat. Nos. 5,262,311, 5,599,672 and 5,965,409; WO97/18454; Prashar and Weissman, *Proc. Natl. Acad. Sci.*,

93:659-663, and U.S. Pat. Nos. 6,010,850 and 5,712,126; Welsh et al., *Nucleic Acid Res.*, 20:4965-4970, 1992, and U.S. Pat. No. 5,487,985) and other RNA fingerprinting techniques, nucleic acid sequence based amplification ("NASBA") and other transcription based amplification systems (e.g., U.S. Pat. Nos. 5,409,818 and 5,554,527; WO 88/10315), polynucleotide arrays (e.g., U.S. Pat. Nos. 5,143,854, 5,424,186; 5,700,637, 5,874,219, and 6,054,270; PCT WO 92/10092; PCT WO 90/15070), Qbeta Replicase (PCT/US87/00880), Strand Displacement Amplification ("SDA"), Repair Chain Reaction ("RCR"), nuclease protection assays, subtraction-based methods, Rapid-Scan™, etc. Additional useful methods include, but are not limited to, e.g., template-based amplification methods, competitive PCR (e.g., U.S. Pat. No. 5,747,251), redox-based assays (e.g., U.S. Pat. No. 5,871,918), Taqman-based assays (e.g., Holland et al., *Proc. Natl. Acad, Sci.*, 88:7276-7280, 1991; U.S. Pat. Nos. 5,210,015 and 5,994,063), real-time fluorescence-based monitoring (e.g., U.S. Pat. 5,928,907), molecular energy transfer labels (e.g., U.S. Pat. Nos. 5,348,853, 5,532,129, 5,565,322, 6,030,787, and 6,117,635; Tyagi and Kramer, *Nature Biotech.*, 14:303-309, 1996). Any method suitable for single cell analysis of gene or protein expression can be used, including in situ hybridization, immunocytochemistry, MACS, FACS, flow cytometry, etc. For single cell assays, expression products can be measured using antibodies, PCR, or other types of nucleic acid amplification (e.g., Brady et al., *Methods Mol. & Cell. Biol.* 2, 17-25, 1990; Eberwine et al., 1992, *Proc. Natl. Acad. Sci.*, 89, 3010-3014, 1992; U.S. Pat. No. 5,723,290). These and other methods can be carried out conventionally, e.g., as described in the mentioned publications.

Many of such methods may require that the polynucleotide is labeled, or comprises a particular nucleotide type useful for detection. The present invention includes such modified polynucleotides that are necessary to carry out such methods. Thus, polynucleotides can be DNA, RNA, DNA:RNA hybrids, PNA, etc., and can comprise any modification or substituent which is effective to achieve detection.

Detection can be desirable for a variety of different purposes, including research, diagnostic, prognostic, and forensic. For diagnostic purposes, it may be desirable to identify the presence or quantity of a polynucleotide sequence in a sample, where the sample is obtained from tissue, cells, body fluids, etc. In a preferred method as described in more detail below, the present invention relates to a method of detecting a polynucleotide

comprising, contacting a target polynucleotide in a test sample with a polynucleotide probe under conditions effective to achieve hybridization between the target and probe; and detecting hybridization.

Any test sample in which it is desired to identify a polynucleotide or polypeptide thereof can be used, including, e.g., blood, urine, saliva, stool (for extracting nucleic acid, see, e.g., U.S. Pat. No. 6,177,251), swabs comprising tissue, biopsied tissue, tissue sections, cultured cells, etc.

Detection can be accomplished in combination with polynucleotide probes for other genes, e.g., genes which are expressed in other disease states, tissues, cells, such as brain, heart, kidney, spleen, thymus, liver, stomach, small intestine, colon, muscle, lung, testis, placenta, pituitary, thyroid, skin, adrenal gland, pancreas, salivary gland, uterus, ovary, prostate gland, peripheral blood cells (T-cells, lymphocytes, etc.), embryo, breast, fat, adult and embryonic stem cells, specific cell-types, such as endothelial, epithelial, myocytes, adipose, etc.

Polynucleotides can be used in wide range of methods and compositions, including for detecting, diagnosing, staging, grading, assessing, prognosticating, etc. diseases and disorders associated with polynucleotides, for monitoring or assessing therapeutic and/or preventative measures, in ordered arrays, etc. Any method of detecting genes and polynucleotides can be used; certainly, the present invention is not to be limited how such methods are implemented.

Along these lines, the present invention relates to methods of detecting polynucleotides in a sample comprising nucleic acid. Such methods can comprise one or more the following steps in any effective order, e.g., contacting said sample with a polynucleotide probe under conditions effective for said probe to hybridize specifically to nucleic acid in said sample, and detecting the presence or absence of probe hybridized to nucleic acid in said sample, wherein said probe is a polynucleotide which is SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, a polynucleotide having, e.g., about 70%, 80%, 85%, 90%, 95%, 99%, or more sequence identity thereto, effective or specific fragments thereof, or complements thereto. The detection method can be applied to any sample, e.g., cultured primary, secondary, or established cell lines, tissue biopsy, blood, urine, stool, cerebral spinal fluid, and other bodily fluids, for any purpose.

Contacting the sample with probe can be carried out by any effective means in any effective environment. It can be accomplished in a solid, liquid, frozen, gaseous, amorphous, solidified, coagulated, colloid, etc., mixtures thereof, matrix. For instance, a probe in an aqueous medium can be contacted with a sample which is also in an aqueous medium, or which is affixed to a solid matrix, or vice-versa.

Generally, as used throughout the specification, the term "effective conditions" means, e.g., the particular milieu in which the desired effect is achieved. Such a milieu, includes, e.g., appropriate buffers, oxidizing agents, reducing agents, pH, co-factors, temperature, ion concentrations, suitable age and/or stage of cell (such as, in particular part of the cell cycle, or at a particular stage where particular genes are being expressed) where cells are being used, culture conditions (including substrate, oxygen, carbon dioxide, etc.). When hybridization is the chosen means of achieving detection, the probe and sample can be combined such that the resulting conditions are functional for said probe to hybridize specifically to nucleic acid in said sample.

The phrase "hybridize specifically" indicates that the hybridization between single-stranded polynucleotides is based on nucleotide sequence complementarity. The effective conditions are selected such that the probe hybridizes to a preselected and/or definite target nucleic acid in the sample. For instance, if detection of a polynucleotide set forth in SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58 is desired, a probe can be selected which can hybridize to such target gene under high stringent conditions, without significant hybridization to other genes in the sample. To detect homologs of a polynucleotide set forth in SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, the effective hybridization conditions can be less stringent, and/or the probe can comprise codon degeneracy, such that a homolog is detected in the sample.

As already mentioned, the methods can be carried out by any effective process, e.g., by Northern blot analysis, polymerase chain reaction (PCR), reverse transcriptase PCR, RACE PCR, *in situ* hybridization, etc., as indicated above. When PCR based techniques are used, two or more probes are generally used. One probe can be specific for a defined sequence which is characteristic of a selective polynucleotide, but the other probe can be specific for the selective polynucleotide, or specific for a more general sequence, e.g., a sequence such as polyA which is characteristic of mRNA, a sequence which is specific for a

promoter, ribosome binding site, or other transcriptional features, a consensus sequence (e.g., representing a functional domain). For the former aspects, 5' and 3' probes (e.g., polyA, Kozak, etc.) are preferred which are capable of specifically hybridizing to the ends of transcripts. When PCR is utilized, the probes can also be referred to as "primers" in that they can prime a DNA polymerase reaction.

In addition to testing for the presence or absence of polynucleotides, the present invention also relates to determining the amounts at which polynucleotides of the present invention are expressed in sample and determining the differential expression of such polynucleotides in samples. Such methods can involve substantially the same steps as described above for presence/absence detection, e.g., contacting with probe, hybridizing, and detecting hybridized probe, but using more quantitative methods and/or comparisons to standards. The amount of hybridization between the probe and target can be determined by any suitable methods, e.g., PCR, RT-PCR, RACE PCR, Northern blot, polynucleotide microarrays, Rapid-Scan, etc., and includes both quantitative and qualitative measurements.

Methods of identifying polymorphisms, mutations, etc., of polynucleotides

Polynucleotides of the present invention can also be utilized to identify mutant alleles, SNPs, gene rearrangements and modifications, and other polymorphisms of the wild-type gene. Mutant alleles, polymorphisms, SNPs, etc., can be identified and isolated from subjects with diseases that are known, or suspected to have, a genetic component. Identification of such genes can be carried out routinely (see, above for more guidance), e.g., using PCR, hybridization techniques, direct sequencing, mismatch reactions (see, e.g., above), RFLP analysis, SSCP (e.g., Orita et al., *Proc. Natl. Acad. Sci.*, 86:2766, 1992), etc., where a polynucleotide having a sequence selected from SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58 is used as a probe. The selected mutant alleles, SNPs, polymorphisms, etc., can be used diagnostically to determine whether a subject has, or is susceptible to a disorder associated with polynucleotides, as well as to design therapies and predict the outcome of the disorder. Methods involve, e.g., diagnosing a disorder associated with polynucleotides or determining susceptibility to a disorder, comprising, detecting the presence of a mutation in a gene represented by a polynucleotide selected from SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58. The detecting can be carried out by any

effective method, e.g., obtaining cells from a subject, determining the gene sequence or structure of a target gene (using, e.g., mRNA, cDNA, genomic DNA, etc), comparing the sequence or structure of the target gene to the structure of the normal gene, whereby a difference in sequence or structure indicates a mutation in the gene in the subject. Polynucleotides can also be used to test for mutations, SNPs, polymorphisms, etc., e.g., using mismatch DNA repair technology as described in U.S. Pat. No. 5,683,877; U.S. Pat. No. 5,656,430; Wu et al., *Proc. Natl. Acad. Sci.*, 89:8779-8783, 1992.

The present invention also relates to methods of detecting polymorphisms in genes of the present invention, comprising, e.g., comparing the structure of: genomic DNA, mRNA, cDNA, etc., with the structure of a polynucleotide set forth in SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, or 58. The methods can be carried out on a sample from any source, e.g., cells, tissues, body fluids, blood, urine, stool, hair, egg, sperm,cerebral spinal fluid, etc.

These methods can be implemented in many different ways. For example, "comparing the structure" steps include, but are not limited to, comparing restriction maps, nucleotide sequences, amino acid sequences, RFLPs, Dnase sites, DNA methylation fingerprints (e.g., U.S. Pat. No. 6,214,556), protein cleavage sites, molecular weights, electrophoretic mobilities, charges, ion mobility, etc.. The term "structure" can refer to any physical characteristics or configurations which can be used to distinguish between nucleic acids and polypeptides. The methods and instruments used to accomplish the comparing step depends upon the physical characteristics which are to be compared. Thus, various techniques are contemplated, including, e.g., sequencing machines (both amino acid and polynucleotide), electrophoresis, mass spectrometer (U.S. Pat. Nos. 6,093,541, 6,002,127), liquid chromatography, HPLC, etc.

To carry out such methods, "all or part" of the gene or polypeptide can be compared. For example, if nucleotide sequencing is utilized, the entire gene can be sequenced, including promoter, introns, and exons, or only parts of it can be sequenced and compared, e.g., exon 1, exon 2, etc.

Mutagenesis

Mutated polynucleotide sequences of the present invention are useful for various purposes, e.g., to create mutations of the polypeptides they encode, to identify functional

regions of genomic DNA, to produce probes for screening libraries, etc. Mutagenesis can be carried out routinely according to any effective method, e.g., oligonucleotide-directed (Smith, M., *Ann. Rev. Genet.*19:423-463, 1985), degenerate oligonucleotide-directed (Hill et al., *Method Enzymology*, 155:558-568, 1987), region-specific (Myers et al., *Science*, 229:242-246, 1985; Derbyshire et al., *Gene*, 46:145, 1986; Ner et al., *DNA*, 7:127, 1988), linker-scanning (McKnight and Kingsbury, *Science*, 217:316-324, 1982), directed using PCR, recursive ensemble mutagenesis (Arkin and Yourvan, *Proc. Natl. Acad. Sci.*, 89:7811-7815, 1992), random mutagenesis (e.g., U.S. Pat. Nos. 5,096,815; 5,198,346; and 5,223,409), site-directed mutagenesis (e.g., Walder et al., *Gene*, 42:133, 1986; Bauer et al., *Gene*, 37:73, 1985; Craik, *Bio Techniques*, January 1985, 12-19; Smith et al., *Genetic Engineering: Principles and Methods*, Plenum Press, 1981), phage display (e.g., Lowman et al., *Biochem.* 30:10832-10837, 1991; Ladner et al., U.S. Pat. No. 5,223,409; Huse, WIPO Publication WO 92/06204), etc. Desired sequences can also be produced by the assembly of target sequences using mutually priming oligonucleotides (Uhlmann, *Gene*, 71:29-40, 1988). For directed mutagenesis methods, analysis of the three-dimensional structure of the polynucleotides polypeptide can be used to guide and facilitate making mutants which effect polypeptide activity. Sites of substrate-enzyme interaction or other biological activities can also be determined by analysis of crystal structure as determined by such techniques as nuclear magnetic resonance, crystallography or photoaffinity labeling. See, for example, de Vos et al., Science 255:306-312, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992.

In addition, libraries of polynucleotides and fragments thereof can be used for screening and selection of polynucleotides variants. For instance, a library of coding sequences can be generated by treating a double-stranded DNA with a nuclease under conditions where the nicking occurs, e.g., only once per molecule, denaturing the double-stranded DNA, renaturing it to for double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single-stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting DNAs into an expression vector. By this method, expression libraries can be made comprising "mutagenized" polynucleotides. The entire coding sequence or parts thereof can be used.

Polynucleotide expression, polypeptides produced thereby, and specific-binding partners thereto.

A polynucleotide according to the present invention can be expressed in a variety of different systems, in vitro and in vivo, according to the desired purpose. For example, a polynucleotide can be inserted into an expression vector, introduced into a desired host, and cultured under conditions effective to achieve expression of a polypeptide coded for by the polynucleotide, to search for specific binding partners. Effective conditions include any culture conditions which are suitable for achieving production of the polypeptide by the host cell, including effective temperatures, pH, medium, additives to the media in which the host cell is cultured (e.g., additives which amplify or induce expression such as butyrate, or methotrexate if the coding polynucleotide is adjacent to a dhfr gene), cycloheximide, cell densities, culture dishes, etc. A polynucleotide can be introduced into the cell by any effective method including, e.g., naked DNA, calcium phosphate precipitation, electroporation, injection, DEAE-Dextran mediated transfection, fusion with liposomes, association with agents which enhance its uptake into cells, viral transfection. A cell into which a polynucleotide of the present invention has been introduced is a transformed host cell. The polynucleotide can be extrachromosomal or integrated into a chromosome(s) of the host cell. It can be stable or transient. An expression vector is selected for its compatibility with the host cell. Host cells include, mammalian cells, e.g., COS, CV1, BHK, CHO, HeLa, LTK, NIH 3T3, cardiac or heart cells, such as W1 (Wang et al., *In vitro Cell. Dev.*, 27:63-74, 1991), MC29, cardiac fibroblasts (e.g., Wang et al., *Tiss Cell.*, 33:86-96, 2001), cardiac microvascular endothelial cells (e.g. Jollow et al., *Transplantation*, 68:430-439, 1999), T/G HA-VSMC (CRL-1999), H9c2(2-1) (CRL-1446), P19 (CRL-1825) , CNS neural stem cells (e.g., U.S. Pat. No. 6,103,530), IMR-32, A172 (ATCC CRL-1620), T98G (ATCC CRL-1690), CCF-STTG1 (ATCC CRL-1718), DBTRG-05MG (ATCC CRL-2020), PFSK-1 (ATCC CRL-2060), SK-N-AS and other SK cell lines (ATCC CRL-2137), CHP-212 (ATCC CRL-2273), RG2 (ATCC CRL-2433), HCN-2 (ATCC CRL-10742), U-87 MG and other U MG cell lines (ATCC HTB-14), D283 Med (ATCC HTB-185), PC12, Neuro-2a (ATCC CCL-131), muscle cells lines, such as RD (CCL-136), G-7, G-8, C2C12, established and primary brain, heart, or muscle cells, G-402 (ATCC CRL-1440), ACHN (ATCC CRL-1611), Vero (ATCC CCL-81), 786-O (ATCC CRL-1932), 769-P (ATCC CRL-1933), CCD 1103

KIDTr (ATCC CRL-2304), CCD 1105 KIDTr (ATCC CRL-2305), Hs 835.T (ATCC CRL-7569), Hs 926.T (ATCC CRL-7678), Caki-1 (ATCC HTB-46), Caki-2 (ATCC HTB-47), SW 839 (ATCC HTB-49), LLC-MK2 (ATCC CCL-7), BHK-21 (ATCC CCL-10), MDBK, CV-1, (ATCC CRL-1573), KNRK (ATCC CRL-1569), NRK-49F (ATCC CRL-1570), A-704 (ATCC HTB-45), and other established and primary kidney lines, CNS neural stem cells (e.g., U.S. Pat. No. 6,103,530), IMR-32, A172 (ATCC CRL-1620), T98G (ATCC CRL-1690), CCF-STTG1 (ATCC CRL-1718), DBTRG-05MG (ATCC CRL-2020), PFSK-1 (ATCC CRL-2060), SK-N-AS and other SK cell lines (ATCC CRL-2137), CHP-212 (ATCC CRL-2273), RG2 (ATCC CRL-2433), HCN-2 (ATCC CRL-10742), U-87 MG and other U MG cell lines (ATCC HTB-14), D283 Med (ATCC HTB-185), PC12, Neuro-2a (ATCC CCL-131), and other established and primary brain cell lines, Hep G2 (ATCC NO. HB-8065), SK-HEP-1 (ATCC NO HTB-52), H2.35 (ATCC NO CRL-1995), CD-1 (ATC NO CRL-2254), C3A (ATCC NO CRL-10741), FL83B (ATCC NO CRL-2390), WRL 68 (ATCC NO CL-48), Hep 3B (ATCC NO HB-8064), insect cells, such as Sf9 (S. frugipeda) and Drosophila, bacteria, such as E. coli, Streptococcus, bacillus, yeast, such as Sacharomyces, S. cerevisiae, fungal cells, plant cells, embryonic or adult stem cells (e.g., mammalian, such as mouse or human).

Expression control sequences are similarly selected for host compatibility and a desired purpose, e.g., high copy number, high amounts, induction, amplification, controlled expression. Other sequences which can be employed include enhancers such as from SV40, CMV, RSV, inducible promoters, cell-type specific elements, or sequences which allow selective or specific cell expression. Promoters that can be used to drive its expression, include, e.g., the endogenous promoter, MMTV, SV40, trp, lac, tac, or T7 promoters for bacterial hosts; or alpha factor, alcohol oxidase, or PGH promoters for yeast. RNA promoters can be used to produced RNA transcripts, such as T7 or SP6. See, e.g., Melton et al., *Polynucleotide Res.*, 12(18):7035-7056, 1984; Dunn and Studier. *J. Mol. Bio.*, 166:477-435, 1984; U.S. Pat. No. 5,891,636; Studier et al., *Gene Expression Technology, Methods in Enzymology,* 85:60-89, 1987. In addition, as discussed above, translational signals (including in-frame insertions) can be included.

When a polynucleotide is expressed as a heterologous gene in a transfected cell line, the gene is introduced into a cell as described above, under effective conditions in which the

gene is expressed. The term "heterologous" means that the gene has been introduced into the cell line by the "hand-of-man." Introduction of a gene into a cell line is discussed above. The transfected (or transformed) cell expressing the gene can be lysed or the cell line can be used intact.

For expression and other purposes, a polynucleotide can contain codons found in a naturally-occurring gene, transcript, or cDNA, for example, e.g., as set forth in SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, or it can contain degenerate codons coding for the same amino acid sequences. For instance, it may be desirable to change the codons in the sequence to optimize the sequence for expression in a desired host. See, e.g., U.S. Pat. Nos. 5,567,600 and 5,567,862.

A polypeptide according to the present invention can be recovered from natural sources, transformed host cells (culture medium or cells) according to the usual methods, Another approach is express the polypeptide recombinantly with an affinity tag (Flag epitope, HA epitope, myc epitope, 6xHis, maltose binding protein, chitinase, etc) and then purify by anti-tag antibody-conjugated affinity chromatography.

The present invention also relates to polypeptides, e.g., an isolated human polypeptide comprising or having the amino acid sequence set forth in SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, an isolated human polypeptide comprising an amino acid sequence having 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more amino acid sequence identity to the amino acid sequence set forth in SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, and optionally having one or more activities. Fragments can also used, e.g., to produce antibodies or other immune responses, as competitors to any activity. These fragments can be referred to as being "specific for" polynucleotides. The latter phrase, as already defined, indicates that the peptides are characteristic of polynucleotides, and that the defined sequences are substantially absent from all other protein types. Such polypeptides can be of any size which is necessary to confer specificity, e.g., 5, 8, 10, 12, 15, 20, etc.

The present invention also relates to specific-binding partners. These include antibodies which are specific for polypeptides encoded by polynucleotides of the present invention, as well as other binding-partners which interact with polynucleotides and polypeptides of the present invention. Protein-protein interactions between polynucleotides

and other polypeptides and binding partners can be identified using any suitable methods, e.g., protein binding assays (e.g., filtration assays, chromatography, etc.) , yeast two-hybrid system (Fields and Song, *Nature*, 340: 245-247, 1989), protein arrays, gel-shift assays, FRET (fluorescence resonance energy transfer) assays, etc. Nucleic acid interactions (e.g., protein-DNA or protein-RNA) can be assessed using gel-shift assays, e.g., as carried out in U.S. Pat. No. 6,333,407 and 5,789,538.

Antibodies, e.g., polyclonal, monoclonal, recombinant, chimeric, humanized, single-chain, Fab, and fragments thereof, can be prepared according to any desired method. See, also, screening recombinant immunoglobulin libraries (e.g., Orlandi et al., *Proc. Natl. Acad. Sci.*, 86:3833-3837, 1989; Huse et al., *Science*, 256:1275-1281, 1989); in vitro stimulation of lymphocyte populations; Winter and Milstein, *Nature*, 349: 293-299, 1991. The antibodies can be IgM, IgG, subtypes, IgG2a, IgG1, etc. Antibodies, and immune responses, can also be generated by administering naked DNA See, e.g., U.S. Pat. Nos. 5,703,055; 5,589,466; 5,580,859. Antibodies can be used from any source, including, goat, rabbit, mouse, chicken (e.g., IgY; see, Duan, W0/029444 for methods of making antibodies in avian hosts, and harvesting the antibodies from the eggs). An antibody specific for a polypeptide means that the antibody recognizes a defined sequence of amino acids within or including the polypeptide. Other specific binding partners include, e.g., aptamers and PNA. Antibodies can be prepared against specific epitopes or domains of polynucleotides.

Any form or type of antibody can be prepared and used. For example, antibodies can be humanized, e.g., where they are to be used therapeutically. Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. The term "antibody" as used herein includes intact molecules as well as fragments thereof, such as Fab, F(ab')2, and Fv which are capable of binding to an epitopic determinant. Such antibody fragments retain some ability to selectively bind with its antigen or receptor. The term "epitope" refers to an antigenic determinant on an antigen to which the paratope of an antibody binds. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific

charge characteristics. Antibodies can be prepared against specific epitopes or polypeptide domains.

Antibodies which bind to polynucleotides polypeptides of the present invention can be prepared using an intact polypeptide or fragments containing small peptides of interest as the immunizing antigen. For example, it may be desirable to produce antibodies that specifically bind to the N- or C-terminal domains of polynucleotides. The polypeptide or peptide used to immunize an animal which is derived from translated cDNA or chemically synthesized which can be conjugated to a carrier protein, if desired. Anti-idiotype technology can also be used to produce invention monoclonal antibodies which mimic an epitope.

Methods of detecting polypeptides

Polypeptides coded for by polynucleotides of the present invention can be detected, visualized, determined, quantitated, etc. according to any effective method. useful methods include, e.g., but are not limited to, immunoassays, RIA (radioimmunassay), ELISA, (enzyme-linked-immunosorbent assay), immunoflourescence, flow cytometry, histology, electron microscopy, light microscopy, in situ assays, immunoprecipitation, Western blot.

Immunoassays may be carried in liquid or on biological support. For instance, a sample (e.g., blood, stool, urine, cells, tissue,cerebral spinal fluid, body fluids, etc.) can be brought in contact with and immobilized onto a solid phase support or carrier such as nitrocellulose, or other solid support that is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled polynucleotides specific antibody. The solid phase support can then be washed with a buffer a second time to remove unbound antibody. The amount of bound label on solid support may then be detected by conventional means.

A "solid phase support or carrier" includes any support capable of binding an antigen, antibody, or other specific binding partner. Supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, and magnetite. A support material can have any structural or physical configuration. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the

surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads

One of the many ways in which gene peptide-specific antibody can be detectably labeled is by linking it to an enzyme and using it in an enzyme immunoassay (EIA). See, e.g., Voller, A., "The Enzyme Linked Immunosorbent Assay (ELISA)," 1978, Diagnostic Horizons 2, 1-7, Microbiological Associates Quarterly Publication, Walkersville, Md.); Voller, A. et al., 1978, J. Clin. Pathol. 31, 507-520; Butler, J. E., 1981, Meth. Enzymol. 73, 482-523; Maggio, E. (ed.), 1980, Enzyme Immunoassay, CRC Press, Boca Raton, Fla.. The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety that can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes that can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, .alpha.-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, .beta.-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods that employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect polynucleotides peptides through the use of a radioimmunoassay (RIA). See, e.g., Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986. The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin,

allophycocyanin, o-phthaldehyde and fluorescamine. The antibody can also be detectably labeled using fluorescence emitting metals such as those in the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Diagnostic

The present invention also relates to methods and compositions for diagnosing a disorder, or determining susceptibility to a disorder, using polynucleotides, polypeptides, and specific-binding partners of the present invention to detect, assess, determine, etc., polynucleotides of the present invention. In such methods, the gene can serve as a marker for the disorder, e.g., where the gene, when mutant, is a direct cause of the disorder; where the gene is affected by another gene(s) which is directly responsible for the disorder, e.g., when the gene is part of the same signaling pathway as the directly responsible gene; and, where the gene is chromosomally linked to the gene(s) directly responsible for the disorder, and segregates with it. Many other situations are possible. To detect, assess, determine, etc., a probe specific for the gene can be employed as described above and below. Any method of detecting and/or assessing the gene can be used, including detecting expression of the gene using polynucleotides, antibodies, or other specific-binding partners.

The present invention relates to methods of diagnosing a disorder associated with a polynucleotide of the present invention, or determining a subject's susceptibility to such

disorder, comprising, e.g., assessing the expression of polynucleotide of the present invention in a tissue sample comprising tissue or cells suspected of having the disorder. The phrase "diagnosing" indicates that it is determined whether the sample has the disorder. A "disorder" means, e.g., any abnormal condition as in a disease or malady. "Determining a subject's susceptibility to a disease or disorder" indicates that the subject is assessed for whether s/he is predisposed to get such a disease or disorder, where the predisposition is indicated by abnormal expression of the gene (e.g., gene mutation, gene expression pattern is not normal, etc.). Predisposition or susceptibility to a disease may result when a such disease is influenced by epigenetic, environmental, etc., factors. Diagnosing includes prenatal screening where samples from the fetus or embryo (e.g., via amniocentesis or CV sampling) are analyzed for the expression of the gene.

By the phrase "assessing expression of a gene or polynucleotide of the present invention," it is meant that the functional status of the gene is evaluated. This includes, but is not limited to, measuring expression levels of said gene, determining the genomic structure of said gene, determining the mRNA structure of transcripts from said gene, or measuring the expression levels of polypeptide coded for by said gene. Thus, the term "assessing expression" includes evaluating the all aspects of the transcriptional and translational machinery of the gene. For instance, if a promoter defect causes, or is suspected of causing, the disorder, then a sample can be evaluated (i.e., "assessed") by looking (e.g., sequencing or restriction mapping) at the promoter sequence in the gene, by detecting transcription products (e.g., RNA), by detecting translation product (e.g., polypeptide). Any measure of whether the gene is functional can be used, including, polypeptide, polynucleotide, and functional assays for the gene's biological activity.

In making the assessment, it can be useful to compare the results to a normal gene, e.g., a gene which is not associated with the disorder. The nature of the comparison can be determined routinely, depending upon how the assessing is accomplished. If, for example, the mRNA levels of a sample is detected, then the mRNA levels of a normal can serve as a comparison, or a gene which is known not to be affected by the disorder. Methods of detecting mRNA are well known, and discussed above, e.g., but not limited to, Northern blot analysis, polymerase chain reaction (PCR), reverse transcriptase PCR, RACE PCR, etc. Similarly, if polypeptide production is used to evaluate the gene, then the polypeptide in a

normal tissue sample can be used as a comparison, or, polypeptide from a different gene whose expression is known not to be affected by the disorder. These are only examples of how such a method could be carried out.

The present invention relates to methods of identifying a genetic basis for a disease or disease-susceptibility, comprising, e.g., determining the association of a disease or disease-susceptibility with a gene of the present invention. An association between a disease or disease-susceptibility and nucleotide sequence includes, e.g., establishing (or finding) a correlation (or relationship) between a DNA marker (e.g., gene, VNTR, polymorphism, EST, etc.) and a particular disease state. Once a relationship is identified, the DNA marker can be utilized in diagnostic tests and as a drug target. Any region of the gene can be used as a source of the DNA marker, exons, introns, intergenic regions, etc.

Human linkage maps can be constructed to establish a relationship between a gene and a disease or condition. Typically, polymorphic molecular markers (e.g., STRP's, SNP's, RFLP's, VNTR's) are identified within the region, linkage and map distance between the markers is then established, and then linkage is established between phenotype and the various individual molecular markers. Maps can be produced for an individual family, selected populations, patient populations, etc. In general, these methods involve identifying a marker associated with the disease (e.g., identifying a polymorphism in a family which is linked to the disease) and then analyzing the surrounding DNA to identity the gene responsible for the phenotype. See, e.g., Kruglyak et al., *Am. J. Hum. Genet.*, 58, 1347-1363, 1996; Matise et al., *Nat. Genet.*, 6(4):384-90, 1994.

Assessing the effects of therapeutic and preventative interventions (e.g., administration of a drug, chemotherapy, radiation, etc.) is a major effort in drug discovery, clinical medicine, and pharmacogenomics. The evaluation of therapeutic and preventative measures, whether experimental or already in clinical use, has broad applicability, e.g., in clinical trials, for monitoring the status of a patient, for analyzing and assessing animal models, and in any scenario involving disease treatment and prevention. Analyzing the expression profiles of polynucleotides of the present invention can be utilized as a parameter by which interventions are judged and measured. Treatment of a disorder can change the expression profile in some manner which is prognostic or indicative of the drug's effect on it. Changes in the profile can indicate, e.g., drug toxicity, return to a normal level, etc.

Accordingly, the present invention also relates to methods of monitoring or assessing a therapeutic or preventative measure (e.g., chemotherapy, radiation, anti-neoplastic drugs, antibodies, etc.) in a subject having a disorder, or, susceptible to such a disorder, comprising, e.g., detecting the expression levels of polynucleotides. A subject can be a cell-based assay system, non-human animal model, human patient, etc. Detecting can be accomplished as described for the methods above and below. By "therapeutic or preventative intervention," it is meant, e.g., a drug administered to a patient, surgery, radiation, chemotherapy, and other measures taken to prevent, treat, or diagnose a disorder. Expression can be assessed in any sample comprising any tissue or cell type, body fluid, etc., as discussed for other methods of the present invention.

The present invention also relates to methods of using polynucleotides binding partners, such as antibodies, to deliver active agents to a tissue for a variety of different purposes, including, e.g., for diagnostic, therapeutic, and research purposes. Methods can involve delivering or administering an active agent to the target tissue, comprising, e.g., administering to a subject in need thereof, an effective amount of an active agent coupled to a binding partner specific for a polypeptide of the present invention, wherein said binding partner is effective to deliver said active agent specifically to the taret tissue.

Any type of active agent can be used in combination with polynucleotides, including, therapeutic, cytotoxic, cytostatic, chemotherapeutic, anti-neoplastic, anti-proliferative, anti-biotic, etc., agents. A chemotherapeutic agent can be, e.g., DNA-interactive agent, alkylating agent, antimetabolite, tubulin-interactive agent, hormonal agent, hydroxyurea, Cisplatin, Cyclophosphamide, Altretamine, Bleomycin, Dactinomycin, Doxorubicin, Etoposide, Teniposide, paclitaxel, cytoxan, 2-methoxycarbonylaminobenzimidazole, Plicamycin, Methotrexate, Fluorouracil, Fluorodeoxyuridin, CB3717, Azacitidine, Floxuridine, Mercapyopurine, 6-Thioguanine, Pentostatin, Cytarabine, Fludarabine, etc. Agents can also be contrast agents useful in imaging technology, e.g., X-ray, CT, CAT, MRI, ultrasound, PET, SPECT, and scintographic.

An active agent can be associated in any manner with a polynucleotides binding partner which is effective to achieve its delivery specifically to the target. Specific delivery or targeting indicates that the agent is provided to the target, without being substantially provided to other tissues. This is useful especially where an agent is toxic, and specific

targeting enables the majority of the toxicity to be aimed at the target tissue, with as small as possible effect on other tissues in the body. The association of the active agent and the binding partner ("coupling") can be direct, e.g., through chemical bonds between the binding partner and the agent, or, via a linking agent, or the association can be less direct, e.g., where the active agent is in a liposome, or other carrier, and the binding partner is associated with the liposome surface. In such case, the binding partner can be oriented in such a way that it is able to bind to polypolypeptide on the cell surface. Methods for delivery of DNA via a cell-surface receptor is described, e.g., in U.S. Pat. No. 6,339,139.

Identifying agent methods

The present invention also relates to methods of identifying agents, and the agents themselves, which modulate polynucleotides and polypeptides of the present invention. These agents can be used to modulate the biological activity of the polypeptide encoded for by the gene, or the gene, itself. Agents which regulate the gene or its product are useful in variety of different environments, including as medicinal agents to treat or prevent disorders associated with polynucleotides of the present invention and as research reagents to modify the function of tissues and cell.

Methods of identifying agents generally comprise steps in which an agent is placed in contact with the gene, its transcription product, its translation product, or other target, and then a determination is performed to assess whether the agent "modulates" the target. The specific method utilized will depend upon a number of factors, including, e.g., the target (i.e., is it the gene or polypeptide encoded by it), the environment (e.g., in vitro or in vivo), the composition of the agent, etc.

For modulating the expression of a gene or polynucleotide, a method can comprise, in any effective order, one or more of the following steps, e.g., contacting a polynucleotide or gene (e.g., in a cell population) with a test agent under conditions effective for said test agent to modulate its expression, and determining whether said test agent modulates it. An agent can modulate expression of polynucleotides at any level, including transcription (e.g., by modulating the promoter), translation, and/or perdurance of the nucleic acid (e.g., degradation, stability, etc.) in the cell.

For modulating the biological activity of a polypeptide, a method can comprise, in any effective order, one or more of the following steps, e.g., contacting a polypeptide (e.g., in a cell, lysate, or isolated) with a test agent under conditions effective for said test agent to modulate the biological activity of said polypeptide, and determining whether said test agent modulates said biological activity.

Contacting polynucleotides with the test agent can be accomplished by any suitable method and/or means that places the agent in a position to functionally control expression or biological activity. Functional control indicates that the agent can exert its physiological effect on polynucleotides through whatever mechanism it works. The choice of the method and/or means can depend upon the nature of the agent and the condition and type of environment in which the polynucleotides is presented, e.g., lysate, isolated, or in a cell population (such as, *in vivo*, *in vitro*, organ explants, etc.). For instance, if the cell population is an *in vitro* cell culture, the agent can be contacted with the cells by adding it directly into the culture medium. If the agent cannot dissolve readily in an aqueous medium, it can be incorporated into liposomes, or another lipophilic carrier, and then administered to the cell culture. Contact can also be facilitated by incorporation of agent with carriers and delivery molecules and complexes, by injection, by infusion, etc.

Agents can be directed to, or targeted to, any part of the polypeptide which is effective for modulating it. For example, agents, such as antibodies and small molecules, can be targeted to cell-surface, exposed, extracellular, ligand binding, functional, etc., domains of the polypeptide. Agents can also be directed to intracellular regions and domains, e.g., regions where the polypeptide couples or interacts with intracellular or intramembrane binding partners.

After the agent has been administered in such a way that it can gain access, it can be determined whether the test agent modulates expression or biological activity. Modulation can be of any type, quality, or quantity, e.g., increase, facilitate, enhance, up-regulate, stimulate, activate, amplify, augment, induce, decrease, down-regulate, diminish, lessen, reduce, etc. The modulatory quantity can also encompass any value, e.g., 1%, 5%, 10%, 50%, 75%, 1-fold, 2-fold, 5-fold, 10-fold, 100-fold, etc. To modulate expression means, e.g., that the test agent has an effect on its expression, e.g., to effect the amount of transcription, to effect RNA splicing, to effect translation of the RNA into polypeptide, to effect RNA or

polypeptide stability, to effect polyadenylation or other processing of the RNA, to effect post-transcriptional or post-translational processing, etc. To modulate biological activity means, e.g., that a functional activity of the polypeptide is changed in comparison to its normal activity in the absence of the agent. This effect includes, increase, decrease, block, inhibit, enhance, etc.

A test agent can be of any molecular composition, e.g., chemical compounds, biomolecules, such as polypeptides, lipids, nucleic acids (e.g., antisense to a polynucleotide sequence selected from SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58), carbohydrates, antibodies, ribozymes, double-stranded RNA, aptamers, etc. For example, if a polypeptide to be modulated is a cell-surface molecule, a test agent can be an antibody that specifically recognizes it and, e.g., causes the polypeptide to be internalized, leading to its down regulation on the surface of the cell. Such an effect does not have to be permanent, but can require the presence of the antibody to continue the down-regulatory effect. Antibodies can also be used to modulate the biological activity of a polypeptide in a lysate or other cell-free form.

Additional cell-based test systems suitable for the analysis of GPCR polypeptides are summarized in Marchese et al. (1999, Trends in Pharmacol. Sci. 20: 370-375) and comprise so-called "ligand screening assays." For example in yeast cells the pheromon receptor can be replaced by a GPCR according to the invention. The effect of test substances on the receptor can be determined upon modulation of histidine synthesis, i.e. by growing in histidine-free medium. In addition using cells transfected with nucleic acids according to the invention it can be analyzed whether test substances mediate translocation of a detectable arrestins, for example of a arrestin-GFP-fusion protein. Moreover, it can be analyzed whether test substances mediate GPCR-mediated dispersion or aggregation of Xenopus laevis melanophores. Another test system utilizes the universal adapter G-protein G alphal6, which mobilizes Ca.sup.2+. Other screening test systems are described in Lemer et al., supra; WO96/41169; U.S. Pat. No. 5,482,835; WO99/06535; EP 0 939 902; WO99/66326; WO98/34948; EP 0 863 214; U.S. Pat. No. 5,882,944 and U.S. Pat. No. 5,891,641.

Therapeutics

Selective polynucleotides, polypeptides, and specific-binding partners thereto, can be

utilized in therapeutic applications. Useful methods include, but are not limited to, immunotherapy (e.g., using specific-binding partners to polypeptides), vaccination (e.g., using a selective polypeptide or a naked DNA encoding such polypeptide), protein or polypeptide replacement therapy, gene therapy (e.g., germ-line correction, antisense), etc.

Various immunotherapeutic approaches can be used. For instance, unlabeled antibody that specifically recognizes a tissue-specific antigen can be used to stimulate the body to destroy or attack a cancer or other diseased tissue, to cause down-regulation, to produce complement-mediated lysis, to inhibit cell growth, etc., of target cells which display the antigen, e.g., analogously to how c-erbB-2 antibodies are used to treat breast cancer. In addition, antibody can be labeled or conjugated to enhance its deleterious effect, e.g., with radionuclides and other energy emitting entitities, toxins, such as ricin, exotoxin A (ETA), and diphtheria, cytotoxic or cytostatic agents, immunomodulators, chemotherapeutic agents, etc. See, e.g., U.S. Pat. No. 6,107,090.

An antibody or other specific-binding partner can be conjugated to a second molecule, such as a cytotoxic agent, and used for targeting the second molecule to a tissue-antigen positive cell (Vitetta, E. S. et al., 1993, Immunotoxin therapy, in DeVita, Jr., V. T. et al., eds, Cancer: Principles and Practice of Oncology, 4th ed., J. B. Lippincott Co., Philadelphia, 2624-2636). Examples of cytotoxic agents include, but are not limited to, antimetabolites, alkylating agents, anthracyclines, antibiotics, anti-mitotic agents, radioisotopes and chemotherapeutic agents. Further examples of cytotoxic agents include, but are not limited to ricin, doxorubicin, daunorubicin, taxol, ethidium bromide, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicine, dihydroxy anthracin dione, actinomycin D, 1-dehydrotestosterone, diptheria toxin, Pseudomonas exotoxin (PE) A, PE40, abrin, elongation factor-2 and glucocorticoid. Techniques for conjugating therapeutic agents to antibodies are well.

In addition to immunotherapy, polynucleotides and polypeptides can be used as targets for non-immunotherapeutic applications, e.g., using compounds which interfere with function, expression (e.g., antisense as a therapeutic agent), assembly, etc. RNA interference can be used in vitro and in vivo to silence polynucleotides when its expression contributes to a disease (but also for other purposes, e.g., to identify the gene's function to change a developmental pathway of a cell, etc.). See, e.g., Sharp and Zamore, *Science*, 287:2431-

2433, 2001; Grishok et al., *Science*, 287:2494, 2001.

Delivery of therapeutic agents can be achieved according to any effective method, including, liposomes, viruses, plasmid vectors, bacterial delivery systems, orally, systemically, etc. Therapeutic agents of the present invention can be administered in any form by any effective route, including, e.g., oral, parenteral, enteral, intraperitoneal, topical, transdermal (e.g., using any standard patch), intravenously, ophthalmic, nasally, local, non-oral, such as aerosal, inhalation, subcutaneous, intramuscular, buccal, sublingual, rectal, vaginal, intra-arterial, and intrathecal, etc. They can be administered alone, or in combination with any ingredient(s), active or inactive.

In addition to therapeutics, *per se*, the present invention also relates to methods of treating a disease showing altered expression of a polynucleotide or polypeptide of the present invention, comprising, e.g., administering to a subject in need thereof a therapeutic agent which is effective for regulating expression of said polynucleotide or polypeptide which is effective in treating said disease. The term "treating" is used conventionally, e.g., the management or care of a subject for the purpose of combating, alleviating, reducing, relieving, improving the condition of, etc., of a disease or disorder. By the phrase "altered expression," it is meant that the disease is associated with a mutation in the gene, or any modification to the gene (or corresponding product) which affects its normal function. Thus, expression of polynucleotides refers to, e.g., transcription, translation, splicing, stability of the mRNA or protein product, activity of the gene product, differential expression, etc.

Any agent which "treats" the disease can be used. Such an agent can be one which regulates the expression of the polynucleotides. Expression refers to the same acts already mentioned, e.g. transcription, translation, splicing, stability of the mRNA or protein product, activity of the gene product, differential expression, etc. For instance, if the condition was a result of a complete deficiency of the gene product, administration of gene product to a patient would be said to treat the disease and regulate the gene's expression. Many other possible situations are possible, e.g., where the gene is aberrantly expressed, and the therapeutic agent regulates the aberrant expression by restoring its normal expression pattern.

Antisense

Antisense polynucleotide (e.g., RNA) can also be prepared from a polynucleotide according to the present invention, preferably an anti-sense to a sequence of SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58. Antisense polynucleotide can be used in various ways, such as to regulate or modulate expression of the polypeptides they encode, e.g., inhibit their expression, for in situ hybridization, for therapeutic purposes, for making targeted mutations (in vivo, triplex, etc.) etc. For guidance on administering and designing anti-sense, see, e.g., U.S. Pat. Nos. 6,200,960, 6,200,807, 6,197,584, 6,190,869, 6,190,661, 6,187,587, 6,168,950, 6,153,595, 6,150,162, 6,133,246, 6,117,847, 6,096,722, 6,087,343, 6,040,296, 6,005,095, 5,998,383, 5,994,230, 5,891,725, 5,885,970, and 5,840,708. An antisense polynucleotides can be operably linked to an expression control sequence. A total length of about 35 bp can be used in cell culture with cationic liposomes to facilitate cellular uptake, but for *in vivo* use, preferably shorter oligonucleotides are administered, e.g. 25 nucleotides.

Antisense polynucleotides can comprise modified, nonnaturally-occurring nucleotides and linkages between the nucleotides (e.g., modification of the phosphate-sugar backbone; methyl phosphonate, phosphorothioate, or phosphorodithioate linkages; and 2'-O-methyl ribose sugar units), e.g., to enhance in vivo or in vitro stability, to confer nuclease resistance, to modulate uptake, to modulate cellular distribution and compartmentalization, etc. Any effective nucleotide or modification can be used, including those already mentioned, as known in the art, etc., e.g., disclosed in U.S. Pat. Nos. 6,133,438; 6,127,533; 6,124,445; 6,121,437; 5,218,103 (e.g., nucleoside thiophosphoramidites); 4,973,679; Sproat et al., "2'-O-Methyloligoribonucleotides: synthesis and applications," Oligonucleotides and Analogs A Practical Approach, Eckstein (ed.), IRL Press, Oxford, 1991, 49-86; Iribarren et al., "2'O-Alkyl Oligoribonucleotides as Antisense Probes," Proc. Natl. Acad. Sci. USA, 1990, 87, 7747-7751; Cotton et al., "2'-O-methyl, 2'-O-ethyl oligoribonucleotides and phosphorothioate oligodeoxyribonucleotides as inhibitors of the in vitro U7 snRNP-dependent mRNA processing event," Nucl. Acids Res., 1991, 19, 2629-2635.

Arrays

The present invention also relates to an ordered array of polynucleotide probes and specific-binding partners (e.g., antibodies) for detecting the expression of polynucleotides or polypeptides in a sample, comprising, e.g., one or more polynucleotide probes or specific

binding partners associated with a solid support or in separate receptacles, wherein each probe is specific for polynucleotides, and the probes comprise a nucleotide sequence of SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58 which is specific for said gene, a nucleotide sequence having sequence identity to SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58 which is specific for said gene or polynucleotide, or complements thereto, or a specific-binding partner which is specific for polynucleotides.

The phrase "ordered array" indicates that the probes (polynucleotides, binding-partners, polypeptides, etc.) are arranged in an identifiable or position-addressable pattern, e.g., such as the arrays disclosed in U.S. Pat. Nos. 6,156,501, 6,077,673, 6,054,270, 5,723,320, 5,700,637, WO09919711, WO00023803. The probes are associated with the solid support in any effective way. For instance, the probes can be bound to the solid support, either by polymerizing the probes on the substrate, or by attaching a probe to the substrate. Association can be, covalent, electrostatic, noncovalent, hydrophobic, hydrophilic, noncovalent, coordination, adsorbed, absorbed, polar, etc. When fibers or hollow filaments are utilized for the array, the probes can fill the hollow orifice, be absorbed into the solid filament, be attached to the surface of the orifice, etc. Probes can be of any effective size, sequence identity, composition, etc., as already discussed.

Transgenic animals

The present invention also relates to transgenic animals comprising polynucleotides genes, and homologs thereof. (Methods of making transgenic animals, and associated recombinant technology, can be accomplished conventionally, e.g., as described in *Transgenic Animal Technology*, Pinkert et al., 2nd Edition, Academic Press, 2002.) Such genes, as discussed in more detail below, include, but are not limited to, functionally-disrupted genes, mutated genes, ectopically or selectively-expressed genes, inducible or regulatable genes, etc. These transgenic animals can be produced according to any suitable technique or method, including homologous recombination, mutagenesis (e.g., ENU, Rathkolb et al., *Exp. Physiol.*, 85(6):635-644, 2000), and the tetracycline-regulated gene expression system (e.g., U.S. Pat. No. 6,242,667). The term "gene" as used herein includes any part of a gene, i.e., regulatory sequences, promoters, enhancers, exons, introns, coding sequences, etc. The polynucleotides nucleic acid present in the construct or transgene can be

naturally-occurring wild-type, polymorphic, or mutated. Where the animal is a non-human animal, its homolog can be used instead. Transgenic animals can be susceptible to any of the dieases and disorders mentioned herein, e.g., as described more particularly under the descriptions of each gene.

Along these lines, polynucleotides of the present invention can be used to create transgenic animals, e.g. a non-human animal, comprising at least one cell whose genome comprises a functional disruption of a polynucleotide of the present invention, or a homolog thereof (e.g., a mouse homolog when a mouse is used). By the phrases "functional disruption" or "functionally disrupted," it is meant that the gene does not express a biologically-active product. It can be substantially deficient in at least one functional activity coded for by the gene. Expression of a polypeptide can be substantially absent, i.e., essentially undetectable amounts are made. However, polypeptide can also be made, but which is deficient in activity, e.g., where only an amino-terminal portion of the gene product is produced.

The transgenic animal can comprise one or more cells. When substantially all its cells contain the engineered gene, it can be referred to as a transgenic animal "whose genome comprises" the engineered gene. This indicates that the endogenous gene loci of the animal has been modified and substantially all cells contain such modification.

Functional disruption of the gene can be accomplished in any effective way, including, e.g., introduction of a stop codon into any part of the coding sequence such that the resulting polypeptide is biologically inactive (e.g., because it lacks a catalytic domain, a ligand binding domain, etc.), introduction of a mutation into a promoter or other regulatory sequence that is effective to turn it off, or reduce transcription of the gene, insertion of an exogenous sequence into the gene which inactivates it (e.g., which disrupts the production of a biologically-active polypeptide or which disrupts the promoter or other transcriptional machinery), deletion of sequences from the gene (or homolog thereof), etc. Examples of transgenic animals having functionally disrupted genes are well known, e.g., as described in U.S. Pat. Nos. 6,239,326, 6,225,525, 6,207,878, 6,194,633, 6,187,992, 6,180,849, 6,177,610, 6,100,445, 6,087,555, 6,080,910, 6,069,297, 6,060,642, 6,028,244, 6,013,858, 5,981,830, 5,866,760, 5,859,314, 5,850,004, 5,817,912, 5,789,654, 5,777,195, and 5,569,824. A transgenic animal which comprises the functional disruption can also be referred to as a

"knock-out" animal, since the biological activity has been "knocked-out." Knock-outs can be homozygous or heterozygous.

For creating functionally disrupted genes, and other gene mutations, homologous recombination technology is of special interest since it allows specific regions of the genome to be targeted. Using homologous recombination methods, genes can be specifically-inactivated, specific mutations can be introduced, and exogenous sequences can be introduced at specific sites. These methods are well known in the art, e.g., as described in the patents above. See, also, Robertson, *Biol. Reproduc.*, 44(2):238-245, 1991. Generally, the genetic engineering is performed in an embryonic stem (ES) cell, or other pluripotent cell line (e.g., adult stem cells, EG cells), and that genetically-modified cell (or nucleus) is used to create a whole organism. Nuclear transfer can be used in combination with homologous recombination technologies.

For example, the polynucleotides locus can be disrupted in mouse ES cells using a positive-negative selection method (e.g., Mansour et al., *Nature*, 336:348-352, 1988). In this method, a targeting vector can be constructed which comprises a part of the gene to be targeted. A selectable marker, such as neomycin resistance genes, can be inserted into a polynucleotides exon present in the targeting vector, disrupting it. When the vector recombines with the ES cell genome, it disrupts the function of the gene. The presence in the cell of the vector can be determined by expression of neomycin resistance. See, e.g., U.S. Pat. No. 6,239,326. Cells having at least one functionally disrupted gene can be used to make chimeric and germline animals, e.g., animals having somatic and/or germ cells comprising the engineered gene. Homozygous knock-out animals can be obtained from breeding heterozygous knock-out animals. See, e.g., U.S. Pat. No. 6,225,525.

The present invention also relates to non-human, transgenic animal whose genome comprises recombinant polynucleotides nucleic acid (and homologs thereof) operatively linked to an expression control sequence effective to express said coding sequence. Such a transgenic animal can also be referred to as a "knock-in" animal since an exogenous gene has been introduced, stably, into its genome.

A recombinant nucleic acid refers to a polynucleotide which has been introduced into a target host cell and optionally modified, such as cells derived from animals, plants, bacteria, yeast, etc. A recombinant nucleic acid includes completely synthetic nucleic acid sequences,

semi-synthetic nucleic acid sequences, sequences derived from natural sources, and chimeras thereof. "Operable linkage" has the meaning used through the specification, i.e., placed in a functional relationship with another nucleic acid. When a gene is operably linked to an expression control sequence, as explained above, it indicates that the gene (e.g., coding sequence) is joined to the expression control sequence (e.g., promoter) in such a way that facilitates transcription and translation of the coding sequence. As described above, the phrase "genome" indicates that the genome of the cell has been modified. In this case, the recombinant polynucleotide has been stably integrated into the genome of the animal. The nucleic acid (e.g., a coding sequence) in operable linkage with the expression control sequence can also be referred to as a construct or transgene.

Any expression control sequence can be used depending on the purpose. For instance, if selective expression is desired, then expression control sequences which limit its expression can be selected. These include, e.g., tissue or cell-specific promoters, introns, enhancers, etc. For various methods of cell and tissue-specific expression, see, e.g., U.S. Pat. Nos. 6,215,040, 6,210,736, and 6,153,427. These also include the endogenous promoter, i.e., the coding sequence can be operably linked to its own promoter. Inducible and regulatable promoters can also be utilized.

The present invention also relates to a transgenic animal which contains a functionally disrupted and a transgene stably integrated into the animal's genome. Such an animal can be constructed using combinations any of the above- and below-mentioned methods. Such animals have any of the aforementioned uses, including permitting the knock-out of the normal gene and its replacement with a mutated gene. Such a transgene can be integrated at the endogenous gene locus so that the functional disruption and "knock-in" are carried out in the same step.

In addition to the methods mentioned above, transgenic animals can be prepared according to known methods, including, e.g., by pronuclear injection of recombinant genes into pronuclei of 1-cell embryos, incorporating an artificial yeast chromosome into embryonic stem cells, gene targeting methods, embryonic stem cell methodology, cloning methods, nuclear transfer methods. See, also, e.g., U.S. Patent Nos. 4,736,866; 4,873,191; 4,873,316; 5,082,779; 5,304,489; 5,174,986; 5,175,384; 5,175,385; 5,221,778; Gordon et al., Proc. Natl. Acad. Sci., 77:7380-7384, 1980; Palmiter et al., Cell, 41:343-345, 1985; Palmiter

et al., Ann. Rev. Genet., 20:465-499, 1986; Askew et al., Mol. Cell. Bio., 13:4115-4124, 1993; Games et al. Nature, 373:523-527, 1995; Valancius and Smithies, Mol. Cell. Bio., 11:1402-1408, 1991; Stacey et al., Mol. Cell. Bio., 14:1009-1016, 1994; Hasty et al., Nature, 350:243-246, 1995; Rubinstein et al., Nucl. Acid Res., 21:2613-2617,1993; Cibelli et al., Science, 280:1256-1258, 1998. For guidance on recombinase excision systems, see, e.g., U.S. Pat. Nos. 5,626,159, 5,527,695, and 5,434,066. See also, Orban, P.C., et al., "Tissue- and Site-Specific DNA Recombination in Transgenic Mice," Proc. Natl. Acad. Sci. USA, 89:6861-6865 (1992); O'Gorman, S., et al., "Recombinase-Mediated Gene Activation and Site-Specific Integration in Mammalian Cells," Science, 251:1351-1355 (1991); Sauer, B., et al., "Cre-stimulated recombination at loxP-Containing DNA sequences placed into the mammalian genome," Polynucleotides Research, 17(1):147-161 (1989); Gagneten, S. et al. (1997) Nucl. Acids Res. 25:3326-3331; Xiao and Weaver (1997) Nucl. Acids Res. 25:2985-2991; Agah, R. et al. (1997) J. Clin. Invest. 100:169-179; Barlow, C. et al. (1997) Nucl. Acids Res. 25:2543-2545; Araki, K. et al. (1997) Nucl. Acids Res. 25:868-872; Mortensen, R. N. et al. (1992) Mol. Cell. Biol. 12:2391-2395 (G418 escalation method); Lakhlani, P. P. et al. (1997) Proc. Natl. Acad. Sci. USA 94:9950-9955 ("hit and run"); Westphal and Leder (1997) Curr. Biol. 7:530-533 (transposon-generated "knock-out" and "knock-in"); Templeton, N. S. et al. (1997) Gene Ther. 4:700-709 (methods for efficient gene targeting, allowing for a high frequency of homologous recombination events, e.g., without selectable markers); PCT International Publication WO 93/22443 (functionally-disrupted).

A polynucleotide according to the present invention can be introduced into any non-human animal, including a non-human mammal, mouse (Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1986), pig (Hammer et al., Nature, 315:343-345, 1985), sheep (Hammer et al., Nature, 315:343-345, 1985), cattle, rat, or primate. See also, e.g., Church, 1987, Trends in Biotech. 5:13-19; Clark et al., Trends in Biotech. 5:20-24, 1987); and DePamphilis et al., BioTechniques, 6:662-680, 1988. Transgenic animals can be produced by the methods described in U.S. Pat. No. 5,994,618, and utilized for any of the utilities described therein.

Database

The present invention also relates to electronic forms of polynucleotides,

polypeptides, etc., of the present invention, including computer-readable medium (e.g., magnetic, optical, etc., stored in any suitable format, such as flat files or hierarchical files) which comprise such sequences, or fragments thereof, e-commerce-related means, etc. Along these lines, the present invention relates to methods of retrieving gene sequences from a computer-readable medium, comprising, one or more of the following steps in any effective order, e.g., selecting a cell or gene expression profile, e.g., a profile that specifies that said gene is expressed in a particular (see the expression profiles described above), and retrieving said differentially expressed gene sequences, where the gene sequences consist of the genes represented by SEQ ID NOS 1, 11, 16, 25, 38, 43, 45, 47, 49, 51, and/or 58, or the polypeptides encoded thereby.

A "gene expression profile" means the list of tissues, cells, etc., in which a defined gene is expressed (i.e, transcribed and/or translated). A "cell expression profile" means the genes which are expressed in the particular cell type. The profile can be a list of the tissues in which the gene is expressed, but can include additional information as well, including level of expression (e.g., a quantity as compared or normalized to a control gene), and information on temporal (e.g., at what point in the cell-cycle or developmental program) and spatial expression. By the phrase "selecting a gene or cell expression profile," it is meant that a user decides what type of gene or cell expression pattern he is interested in retrieving, Any pattern of expression preferences may be selected. The selecting can be performed by any effective method. In general, "selecting" refers to the process in which a user forms a query that is used to search a database of gene expression profiles. The step of retrieving involves searching for results in a database that correspond to the query set forth in the selecting step. Any suitable algorithm can be utilized to perform the search query, including algorithms that look for matches, or that perform optimization between query and data. The database is information that has been stored in an appropriate storage medium, having a suitable computer-readable format. Once results are retrieved, they can be displayed in any suitable format, such as HTML A query is formed by the user to retrieve the set of genes from the database having the desired property. Once the query is inputted into the system, a search algorithm is used to interrogate the database, and retrieve results.

Advertising, licensing, etc., methods

The present invention also relates to methods of advertising, licensing, selling, purchasing, brokering, etc., genes, polynucleotides, specific-binding partners, antibodies, etc., of the present invention. Methods can comprises, e.g., displaying a gene, polynucleotide, polypeptide, or antibody specific for a polypeptide in a printed or computer-readable medium (e.g., on the Web or Internet), accepting an offer to purchase said gene, polypeptide, or antibody, etc.

Other

A polynucleotide, probe, polypeptide, antibody, specific-binding partner, etc., according to the present invention can be isolated. The term "isolated" means that the material is in a form in which it is not found in its original environment or in nature, e.g., more concentrated, more purified, separated from component, etc. An isolated polynucleotide includes, e.g., a polynucleotide having the sequenced separated from the chromosomal DNA found in a living animal, e.g., as the complete gene, a transcript, or a cDNA. This polynucleotide can be part of a vector or inserted into a chromosome (by specific gene-targeting or by random integration at a position other than its normal position) and still be isolated in that it is not in a form that is found in its natural environment. A polynucleotide, polypeptide, etc., of the present invention can also be substantially purified. By substantially purified, it is meant that polynucleotide or polypeptide is separated and is essentially free from other polynucleotides or polypeptides, i.e., the polynucleotide or polypeptide is the primary and active constituent. A polynucleotide can also be a recombinant molecule. By "recombinant," it is meant that the polynucleotide is an arrangement or form which does not occur in nature. For instance, a recombinant molecule comprising a promoter sequence would not encompass the naturally-occurring gene, but would include the promoter operably linked to a coding sequence not associated with it in nature, e.g., a reporter gene, or a truncation of the normal coding sequence.

The term "marker" is used herein to indicate a means for detecting or labeling a target. A marker can be a polynucleotide (usually referred to as a "probe"), polypeptide (e.g., an antibody conjugated to a detectable label), PNA, or any effective material.

The topic headings set forth above are meant as guidance where certain information can be found in the application, but are not intended to be the only source in the application where information on such topic can be found. Reference materials

For other aspects of the polynucleotides, reference is made to standard textbooks of molecular biology. See, e.g., Hames et al., Polynucleotide Hybridization, IL Press, 1985; Davis et al., Basic Methods in Molecular Biology, Elsevir Sciences Publishing, Inc., New York, 1986; Sambrook et al., Molecular Cloning, CSH Press, 1989; Howe, Gene Cloning and Manipulation, Cambridge University Press, 1995; Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, Inc., 1994-1998.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever. The entire disclosure of all applications, patents and publications, cited above and in the figures are hereby incorporated by reference in their entirety, including U.S. Serial No. 10/094,092, filed March 11, 2002, U.S. Serial No. 10/112,372, filed April 1, 2002, U.S. Serial No. 60/382, 614, filed May 24, 2002, U.S. Serial No. 10/164,717, filed June 10, 2002, U.S. Serial No. 10/167,631, filed June 13, 2002, U.S. Serial No. 10/177,917, filed June 24, 2002, and U.S. Serial No. 60/399,125, filed 30 July 2002, which are hereby incorporated by reference in their entirety.

TABLE 1

Variant[a]	Nucleotide change
Pro18Ala	52C > G
86insA	
Val60Leu	178G > T
Ala64Ser	190G > T
Arg67Gln	200G > A
Phe76Tyr	227T > A
Asp84Glu	252C > A
Ala81Pro	241G > C
Val92Met	274G > A
Thr95Met	284C > T
Val97Ile	289G > A
Ala103Val	308C > T
Gly104Ser	310G > A
Leu106Gln	317T > A
Leu106Leu	318G > A
Arg142His	425G > A
Arg151Cys	451C > T
Arg151Arg	453C > G
Ile155Thr	464T > C
Arg160Trp	478C > T
Arg163Gln	488G > A
Val173del	
Val174Ile	520G > A
537insC	
Pro230Leu	689C > T
Pro230Pro	690G > A
Gln233Gln	699G > A
His260Pro	779A > C
Ile264Ile	792C > T
Cys273Cys	819C > T
Lys278Glu	832A > G
Asn279Ser	836A > G
Asn279Lys	837C > A
Ile287Met	861C > G
Asp294His	880G > C
Phe300Phe	900C > T
Thr314Thr	942A > G
Ser316Ser	948C > T

TABLE 2

Allele	Allele Frequency, %		Stimulation of cAMP Production
	White Populations	Individuals With Red Hair	
Wild type	53	23	+++
Val60Leu†	10	3	+
Ala64Ser	<1	1	NA
Lys65Asn	<1	<1	NA
Arg67Gln	0‡	0	NA
Arg67Val	0‡	0	NA
Phe76Tyr	<1	<1	NA
Asp84Glu	1	3	+++
Asn91Asp	<1	0	NA
Val92Leu	<1	1	NA
Val92Met	8	8	+++
Thr95Met	<1	1	NA
Val97Ile	<1	<1	NA
Ala103Val	<1	<1	NA
Leu106Gln	<1	<1	NA
Arg142His	<1	1	-
Arg151Cys§	8	25	-
Ile155Thr	<1	<1	NA
Arg160Trp§	7	19	-
Arg163Gln	4‖	<1	NA
Ile287Met	0‡	0	NA
Asp294His§	4	13	-
Ala299Thr	<1	<1	NA
ins29¶	<1	<1	-
ins179¶	<1	<1	-

*Several synonymous variants have also been described, including Leu106Leu, Leu158Leu, Gln233Gln, Cys273Cys, Phe300Phe, Thr314Thr, and Ser316Ser. MC1-R indicates melanocortin-1 receptor; cAMP, cyclic adenosine monophosphate; triple plus sign, significant stimulation (same as wild type); single plus sign, minimal stimulation; NA, data not available; and minus sign, no stimulation (nonfunctional receptor).

†Possible association with blond/fair hair.

‡Present in <1% of East/Southeast Asians.

§Strong association with red hair, fair skin, and poor tanning ability; recent work also shows an association with cutaneous melanoma and nonmelanoma skin cancer.

‖Present in >70% of East/Southeast Asian and Native Americans.

¶ins indicates insertion; these single-nucleotide insertion mutations produce frameshifts that result in a prematurely terminated, nonfunctioning